THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Mehdi Senoussi

Le 5 février 2016

Titre :

Temporal and spatial flexibility of neural representations of visual objects through learning

Flexibilité temporelle et spatiale des représentations neurales d'objets visuels lors d'apprentissages

---

École doctorale et discipline ou spécialité :

ED CLESCO : Neurosciences, comportement et cognition

Unité de recherche :

CerCo - CNRS UMR 5549 - Université Paul Sabatier, Toulouse 3

Directeur(s) de Thèse :

Rufin VanRullen
Leila Reddy

Rapporteurs :

Floris de Lange
Olivier Pascalis

«...I asked Mel [Melvyn Goodale], that he - like Sperry and Eccles - might succumb to that ill-defined mysticism that often overtakes neuroscientists in late career? And what form might this take? With a certain inevitability, we cast an argument in terms of two visual systems. It seemed to us both that, if pushed, God might well wish to exert his influence on the world through the *dorsal* stream, where his manipulations would go unnoticed by consciousness. What then of the ventral stream, I asked? With mock seriousness, Mel chose to display his contempt for those who have viewed vision solely as a vehicle for passive perception, responding: "The ventral stream is the Devil's playground"! »

Turnbull, Oliver H. "Of two minds about two visual systems." *Psyche* 5, no. 8 (1999): 1-5.

Acknowledgments

This PhD has been a long and life-changing journey. Besides the actual job of learning to be a scientist, about methodologies, techniques and the state of the art, etc., I have learned many things: why I pursued a scientific career, what are my strengths and weaknesses, in sum, who I was and who I wanted to be (at least to some extent).

I know for a fact that without the people who supported me during this PhD and the years that led to it, I would not have been able to do it. So I would like to acknowledge the help, guidance and support from quite a few people and institutions which played a crucial role in this chapter of my life.

First I would like to thank my supervisors: Leila Reddy and Rufin VanRullen. Thank you for your constant support throughout this journey. Thank you for giving me the chance to step into the research world. No matter what, you have always been available for me, you helped me get on with things when I was getting lost in the meanderings of research (or of my own mind), always with a pragmatic point of view and a clear plan. I realize how lucky I am to have had such a constant presence even though you had so many other things to do. Scientifically I can not begin to express how lucky I am to have had supervisors who were able to guide me from the theoretical to the most technical level. You have taught me so many things from the work ethic, to the methodology, organization skills, critical thinking, etc. Again, it has been a long trip but I'm grateful to have been guided by you.

I would like to thank my reviewers Floris de Lange and Olivier Pascalis for accepting to review my manuscript and evaluate my defense. I thank you for your critical and thoughtful reviews, your challenging and stimulating questions and comments. I look forward to future opportunities to interact with you.

I would like to acknowledge the people at the CerCo lab, and more specifically the people who made what this lab is now starting with Michele Fabre-Thorpe and Simon Thorpe who have created such an nice and stimulating environment to work in, both with a firm management and rules that structured the lab and brought us all together, for giving us the opportunity to meet brilliant invited speakers, but also with all the more relaxed occasions such as winter/summer retreats, Christmas 'parties', etc. Also Claire Thoueilles for being so helpful and to have made my administrative life so much easier. Thanks also to SITH for their support and especially to Maxime for the python and computer related discussions and advises.
I would like to thank the MRI platform staff, especially Nathalie Vayssière, for helping us set up the experiment and the technical support. Pour finir je voudrais remercier Maroufa et Fatma pour leur constante bonne humeur, leur sourire chaleureux, vous êtes, comme Edd l'a si bien dit, un rayon de soleil au labo.

These 3 years would not have been as fun and interesting without my colleagues in the PAF and MAOS teams. The first person I met was Laura who welcomed me in the lab and who was always there for advices, for listening to my grievances and believing in me. Also for the friday dance-offs in our office. I'm so grateful I met the brilliant scientist and wonderful friend that you are. I also want to thank Douglas McLelland and Sebastien Crouzet, the "Super-postdocs", for your humility, wisdom,

and humor. I really wish there were more people like you in research. And to all the past and present phds and postdocs from these teams: Anne-Claire, Tracy, Grace, Biao, Marie, Marina, Rodika, Adrien, Gab, Sasskia, thank you the interesting discussions and the fun times both in and outside the lab. Lola for your constant optimism, ability to put things in perspective, the deep, philosophical discussions about research (and hip-hop) and for being the only other "*parigo*" with me in this lab. Thomas for being far more than a colleague, for the NTMs, for the roof, for all the discussions, thank you. To Edd for making me question cultural differences over and over and over again, for your the insightful reflexions, for your sensibility and for being here and yourself my friend.

I would also like to thank Thérèse Collins, my past supervisor, who has been present as a mentor from my masters and always took the time to give me wise and insightful thoughts on my career. And to the entire CogMaster organization for creating such an incredibly diverse community of students with the same passion for cognition. To my fellow ex-cogmasterians, Marianne, Alizée, Antoine, Auréliane, Marwa, Manu, for the inspiring discussions about science, amongst other things, and for coming to my big day. To Juliane for having been a huge support during the first stages of my PhD.

The support also (and maybe most importantly) came from outside the lab starting with my roommates Miguel, Bea, thanks for taking my mind out of the dissertation writing stress and always being positive. Flora for being like a sister and welcoming me into your life and friends. Matthieu for being such a caring and thoughtful friend. And Margot for being the most selfless person I know, for having listened to my grievances e-ve-ry-day for the past 3 years and for looking after me and being the best cook I know :).
To Aurélie who has gone from the coolest person in the InCOGnu association to the coolest person on earth. Thank you for your unshakeable support, for believing in me and for knowing how to make me laugh and look up even when I feel down.

To the ones who have been here way before I started this PhD. Pierre for still being here after all these years, even though you've waited a lot, for putting things in perspective, for your wisdom and pragmatism and for being present no matter what when I needed you. To Raf, for being so caring and selfless.

A mon père pour m'avoir insufflé cette inétanchable soif d'apprendre et de comprendre, d'avoir été si positif, encourageant et généreux. A ma mère pour nous avoir donné tous les moyens de réussir ce que nous entreprenions, nous avoir poussé à travailler dur et d'être un exemple de persévérance et de courage. A mon frère Réda pour avoir cru en moi et m'avoir donné la confiance et la détermination d'accomplir tout ce que je souhaitais et d'avoir été mon modèle pendant tant d'années. A ma sœur Amel pour avoir stimulé ma curiosité et mon esprit scientifique sans même s'en rendre compte. A ma sœur Houria et tata Tamina pour leur confiance en moi et leur soutient. A la délégation Midi-Pyrénées de la famille Yasmina, Philippe, Yanis et Kamélia pour leur soutient et leur bonne humeur.

Enfin à toi qui lis encore ces lignes, peut être en espérant y trouver ton nom...j'espère que tu voudras bien pardonner mon oubli.

Table of contents

Abstract (english)

Natural scene processing lies at the very core of everyday vision. The visual system is constantly engaged in the processing of natural objects and scenes, yet much of vision science is conducted with artificial, simple stimuli such as oriented bars and gratings. However, the mechanisms underlying the processing of these simpler stimuli do not necessarily carry over to natural scene processing natural scenes appear to be “special” for the visual system. However, the influence of visual experience can occur with many different types of stimuli, and on multiple different timescales: evolutionary (e.g., with faces, animals or outdoor scenes), lifetime (e.g., with man-made objects such as vehicles or tools, or language characters during acquisition of a native language), or in the shorter-term, over a few months or years (e.g., in a lab setting). These different scales of expertise may potentially have markedly different effects on behavioral and neural processes. In this project we investigated the effects of visual learning on behavior and object representations as a function of the timescale of experience. We conducted two experiments in short and long-term settings and showed that the temporal as well as spatial components of visual processing could be altered, modified by experience. We first showed that when learning a sequence of visual stimuli the brain generates spontaneous selective patterns of activity in the absence of the expected event, probably in order to prepare itself to perceive faster and with better accuracy what surrounds us. This study was conducted with EEG recording and permitted us to show that these mechanisms are generated in the high-alpha and low-beta frequency bands. We then conducted experiment on the flexibility of existing neural representations of objects. In this experiment we made subjects learn associations between faces, cars, houses and chairs and measured how much a 3-weeks training would impact the neural representations of these categories using fMRI. We showed that there was a global shift of 3.3% in these representations and, when focusing on certain function areas of the ventral visual pathway, we observed shifts until 5.5%, which were highly significant ($p<10-4$). We were thus able to show that learning associations between visual stimuli in a short or long-term protocol induces changes in the temporal and spatial dimensions of the visual system. The work conducted in this thesis provides insight in learning processes and can help us understand when learning is impaired (in dyslexia for example) or general memory impairments as well as designing artificial systems that can learn.

Abstract (français)

Le système visuel est constamment engagé dans le traitement des objets naturels et des scènes, mais une grande partie de la recherche dans ce domaine est menée à l'aide de stimuli artificiels et simples tels que les barres orientés. Cependant, les mécanismes sous-jacents au traitement de ces stimuli simples ne se généralisent pas nécessairement à la perception de scènes naturelles. L'influence de l'expérience visuelle peut se produire avec de nombreux types de stimuli, et sur des échelles de temps différentes: évolutionnaire (par exemple, avec des visages ou animaux), la durée d'une vie (par exemple, avec des objets fabriqués par l'homme, ou les caractères écrits d'une langue), ou dans le court terme, sur quelques mois ou années (par exemple, dans un laboratoire). Dans ce projet, nous avons étudié les effets de l'apprentissage visuel sur le comportement et les représentations neurales d'objets en fonction de l'échelle de temps de l'expérience. Nous avons effectué deux expériences dans des contextes à court et à long. Nous avons d'abord montré que l'apprentissage d'une séquence de stimuli visuels du cerveau génère motifs sélectifs spontanées d'activité en l'absence de l'événement attendu, probablement dans le but de se préparer à percevoir plus rapidement et avec plus de précision ce qui nous entoure. Cette étude a été menée avec l'enregistrement EEG et nous a permis de montrer que ces mécanismes sont générés dans les bandes de fréquences à faible bêta haute-alpha et. Nous avons ensuite réalisé l'expérience sur la flexibilité des représentations neuronales existants d'objets. Dans cette expérience, nous avons fait les sujets apprennent associations entre des visages, des voitures, des maisons et des chaises et mesuré combien une formation de 3 semaines aurait un impact sur les représentations neuronales de ces catégories à l'aide IRMf. Nous avons montré qu'il y avait une réorganisation globale dans ces représentations, et lorsque cette analyse se concentre sur certaines zones fonctionnelles voie visuelle ventrale, nous avons observé des changements hautement significatifs, jusqu'à 5,5% ($p < 10^{-4}$). Nous avons ainsi pu montrer que les associations d'apprentissage entre des stimuli visuels dans un protocole à court et à long terme induisent des changements dans les traitement temporel et spatial du système visuel. Les travaux menés dans cette thèse permettent de mieux comprendre les processus d'apprentissage visuels et peuvent nous aider à comprendre quand l'apprentissage est altéré (dans la dyslexie par exemple) ou dans certains troubles de la mémoire plus généralement.

Résumé substantiel

Nos capacités de reconnaissance visuelle d'objets ne semblent pas nécessiter d'efforts ou de temps en règle générale. Malgré leur apparente automaticité elles sont le produit d'un long apprentissage, sur des années pour apprendre à lire par exemple. La façon dont le système visuel est modelé par ces nouveaux apprentissages et en intègre constamment de nouveaux afin de pouvoir reconnaitre des objets et interagir avec eux de façon adaptée reste méconnue. Les études séminales d'Ungerleider et Mishkin ont mis en évidence le rôle de la voie occipito-temporale ventrale dans nos fonctions de reconnaissances d'objets. Des lésions localisées de cette voie produisent des agnosies visuelles : une incapacité à reconnaitre certains objets. A l'aide de technique de neuro-imagerie comme l'électro-encéphalographie (EEG) ou l'Imagerie par Résonance Magnétique (IRM) il est maintenant possible d'étudier les représentations neurales sous-tendant notre perception visuelle, en observant la structure de l'activité enregistrée (ou patterns d'activations), élicitée par la présentation de différentes catégories d'objets.

De quelle façon nos apprentissages façonnent-ils la dynamique temporelle des traitements de notre système visuel et la structure spatiale de ces représentations neurales ?

Effets d'un apprentissage sur la dynamique temporelle des traitements neuraux. (*Article en préparation*)

Afin d'étudier l'effet d'un apprentissage à court terme sur la dynamique des traitements neuraux de catégories visuelles nous avons réalisé une expérience d'apprentissage d'une séquence de six images (visage, maison, voiture, chaise, chameau, pomme de pin) où les sujets avaient pour instruction d'apprendre l'ordre de présentation des images. 16 sujets ont été recrutés et leur activité EEG enregistrée pendant qu'ils apprenaient cette séquence. Deux types d'essais étaient aléatoirement présentés : soit l'image apparaissait, soit elle était remplacée par un carré gris, les essais "catch". Notre hypothèse était que lors ces essais l'activité neurale enregistrée refléterait la catégorie du stimulus qui aurait du apparaitre. En d'autres termes, la préparation du

système visuel à l'apparition prévu d'un stimulus évoquerait une activité sélective représentant ce stimulus.

A l'aide de techniques de classification (Support Vector Machine, SVM) nous avons pu montrer que l'activité évoquée lors des essais "catch" était différente en fonction de l'image qui aurait du apparaitre. Cette activité devenait sélective aussi tôt que 120ms après l'apparition prévue du stimulus attendu. Ce premier résultat nous pousse à penser que cette activité sélective est spontanée et le produit d'un mécanisme automatique qui traque les régularités dans l'environnement afin de faire des prédictions et de préparer les systèmes sensoriels au mieux à détecter des choses importantes dans le monde extérieur.

Le second résultat obtenu est que cette activité sélective était présente dans les électrodes frontales et, dans une moindre mesure, occipito-pariétales, ce qui suggère que les résultat précoces de sélectivité de la distribution des potentiels évoqués provient d'un phénomène de contrôle provenant du cortex frontal et/ou de l'hippocampe qui activerait sélectivement les représentations des stimuli qu'il prévoit de voir. De plus une analyse en classification croisée à permis de montrer que des patterns d'activité similaire étaient évoqués lors des essais Stim et Catch. Ce résultat suggère qu'il y aurait une évocation des représentations neurales sous-tendant la perception. Ce résultat rappelle aussi les recherches menées sur l'imagerie mentale qui ont montrées que lorsque nous générons mentalement une image que nous connaissons les substrats neuronaux qui nous permettent de percevoir cette image sont activés et activés de la même façon que lorsque nous la percevons, c'est à dire que le code neural est partagé entre perception et imagerie. Il serait donc intéressant à partir de la de mené une nouvelle expérience qui inclurait un bloc d'essais d'imagerie mentale afin de comparer l'activité évoquée lors des essais Catch et celle évoquée par la production volontaire d'images mentales.

Afin de mieux caractériser les mécanismes ayant donnés naissances à cette activité sélective nous avons mené une analyse temps-fréquence qui permet de décomposer le signal en bandes de fréquences afin d'étudier la force et l'informativité de chacune d'entre elle car il a été montré par le passé que les mécanismes neuraux de la mémoire et de l'attention, entre autres, sont sous-tendus par certaines bandes de

fréquences (Theta, Alpha et Beta). Notre troisième résultat a donc été de montrer que cette activité sélective se manifeste dans les bandes de fréquences Alpha et Beta du spectre d'activité EEG. La présence d'information dans la bande beta rappelle un résultat récent où il a été montré que la bande beta était privilégié pour les informations de retour (feedback), ce qui semble cohérent avec notre interprétation qui est que ce sont des phénomènes de haut niveau qui élicitent

Ces résultats préliminaires montrent qu'une activité sélective apparait lors des essais "catch" à partir de 120ms après l'apparition du carré gris. L'apprentissage de cette séquence d'images a induit l'activation spontanée de la représentation d'un stimulus dans un contexte ou celui-ci est attendu, possiblement afin de se préparer à sa perception ou sa détection, donnant ainsi un avantage évolutif à ce phénomène neural.

Effet d'un apprentissage à long-terme sur la structure des représentations neurales d'objets (*article soumis*)

La flexibilité du système occipito-temporal visuel a été grandement étudiée depuis des décennies, montrant des capacités d'intégration de l'information lors d'apprentissages à reconnaître ou discriminer de nouveaux objets. La lecture est un exemple idéal et de nouvelles avancées ont aidé à mieux comprendre son fonctionnement en enregistrant l'activité cérébrale de zones impliquées dans la reconnaissance de la forme visuelle des lettres et d'autres objets chez des sujets lettrés et illettrés. Ces études ont permis de montrer que les régions fonctionnelles sensibles à différents types d'objets (visages, objets animés, objets inanimés, ect.) allait subir une restructuration dans leur topographie fonctionnelle et même une latéralisation partielle des substrats neuraux du traitement visuel des visages afin, peut-être, de laisser une place au traitement de la forme visuelle des mots.

Mais la façon dont de nouveaux apprentissages impliquant des catégories déjà connues **s'intègrent** à long terme dans **l'organisation neurale déjà établie** des représentations visuelles d'objets est peu connue. Pour étudier quels changements un apprentissage à long terme induirait sur ces représentations neurales nous avons recrutés 20 participants qui ont appris durant 3 semaines des associations entre

catégories d'images (par exemple : visages<->voitures). Les représentations neurales des différentes catégories d'objets ont été enregistrées avant et après cet apprentissage afin de mesurer les changements que celui-ci entrainerait dans l'organisation de ces représentations.

Nous avons pu mettre en évidence que les représentations de catégories d'objets associées avaient changé, elles étaient devenues plus similaires l'une de l'autre, du à cette apprentissage associatif. Il y a donc une réorganisation possible des représentations neurales visuelles de catégories d'objets. Nous avons ensuite testé les effets de cette réorganisation sur la perception de ces catégories dans une tâche d'amorçage. Les résultats de cette expérience complémentaire montrent que cet apprentissage a produit un effet de facilitation des temps de réaction dans la détection d'une catégorie lorsque celle associée était présentée en amorce. De plus, pour chaque sujet, l'effet de facilitation perceptif mesuré était corrélé avec la magnitude de l'effet de réorganisation des représentations neurales observé en IRM.

Ces résultats montrent que l'intégration d'apprentissages est possible même dans la structure des représentations neurales de catégories visuelles d'objets déjà établies et que cette réorganisation affecte le traitement perceptif de ces objets.
De plus le pic de modification de ces représentations était localisé dans le gyrus fusiforme, une région du cortex impliquée notamment dans l'apprentissage de la lecture. Ce dernier résultat montre une nouvelle fois une plasticité particulière de cette région du cerveau aussi impliquée dans l'expertise d'objets visuels tel qu'acquiert les ornithologue pour les oiseaux ou les passionnés des voitures.

Conclusion :

L'utilisation de l'IRM et de l'EEG nous a permis d'étudier les effets d'apprentissages à différentes échelles de temps et sur différentes dynamiques neurales : temporelle et spatiale. Ces deux études montrent des capacités de réorganisation de notre système visuel, une flexibilité dans sa dynamique temporelle et de son organisation spatiale pour intégrer de nouvelles informations. Comprendre comment de nouveaux apprentissages s'intègrent dans la structure de notre système visuel nous permettra de mieux comprendre pourquoi, dans certains cas, cette intégration ne se fait pas,

dans certains troubles de l'apprentissage, ou quelles étapes des cette intégration sont lésées dans certaines maladies affectant la mémoire.

Chapter I: Introduction

I. General introduction

Understanding how the brain handles visual information that enters our retina to make sense of the world is one of the oldest inquiries in cognitive sciences. Rightly so, the visual system is the most developed of our senses by the spatial extent of the cortex devoted to it or how much we rely on this specific sensory information for much of our perceptually guided behavior. It is therefore a particularly relevant system to study in a *broader* attempt to understand neural implementation underlying general cognitive mechanisms.

Our visual abilities are often regarded as incredible feats of efficiency in terms of stability and flexibility. Stability because we are able to recognize and interact with objects despite changes in viewing conditions (pose, illumination, etc.). Flexibility because depending on the task at hand the relevant **information** can be at the level of category of the object ("is it a mug?"), its identity ("is it my mug?"), estimation of the object's function ("how to pick it up?"), etc.

When examining the scope of possible stimuli we process automatically and effortlessly, we are faced with a problem on the origin and nature of the processes that allow us to interact with the world through our visual inputs. Some "objects" such as fruits or vegetables, animals, spatial layouts (mountains, forests, etc.) have been around us for a period necessary to exert an evolutionary pressure on our visual system to make it efficient at perceiving them. But when it comes to more recent classes such as letters or certain tools we are still able to recognize and interact with them at a similar speed and accuracy (VanRullen and Thorpe, 2001). Our visual system is thus able to learn, it is shaped by experience, to some extent at least, in order to be able to operate efficiently on objects which became relevant only a few centuries or even decades ago. How is it able to adapt, to show flexibility while keeping stability? How does it integrate new knowledge in an already established organization?

To investigate how our visual system learns we will first present the basis of its neural implementation. We will then review the advances in the study of neural

object representations. Finally we will summarize what is known from memory and learning systems behaviorally and the neural substrates underlying these processes.

A. Visual functions in the brain

As often, trying to understand how a system works is helped by studying cases when it does not. More than a century ago, reports of visual impairments due to cortical lesions have lead the way in linking visual functions to their neural substrates.

1. Impairments of visual faculties: clinical cases

a. Optic ataxia

One of the first reports of visually guided behavior impairment came from Rezso Bàlint in 1909 in a medical report of a patient who, after suffering lesions in his posterior parietal lobes, was unable to reach accurately for objects (Rossetti et al., 2003). This patient could nevertheless make precise reaches to different parts of his body with eyes closed, ruling out a strictly motor impairment. This account was one of the first links established between cerebral lesion and visuo-motor ability; Bàlint named it "*Optische Ataxie"* (optic ataxia) meaning a lack of voluntary visual coordination of muscle movements.

Later on Ungerleider and Mishkin showed in a seminal study that lesions in monkey's parietal cortex led to impairments of landmark discrimination (Mishkin and Ungerleider, 1982). Monkeys had to choose one of two covered foodwell which was closest to a cylinder that could be moved across the table. Bilateral removal of the posterior parietal cortex led to severe impairments in this task, proving a critical role of the parietal cortex in spatial and reaching abilities. They thus termed it the "where" pathway.

More recently this theory was tested in a patient who suffered bilateral lesions of the occipito-parietal cortex following strokes (Goodale et al., 1994). Using a clever grasping paradigm they were able to show that when instructed to pick up objects of different shapes using her thumb and index, controls positioned their finger so

that the line passing by the two contact points on the object would be near its center of mass, and thus make the lifting easy, see fig 1. RV, despite her ability to describe these shapes, consistently chose unstable grasp points, the line between the contact points of the two fingers was off of the center of mass of the object.

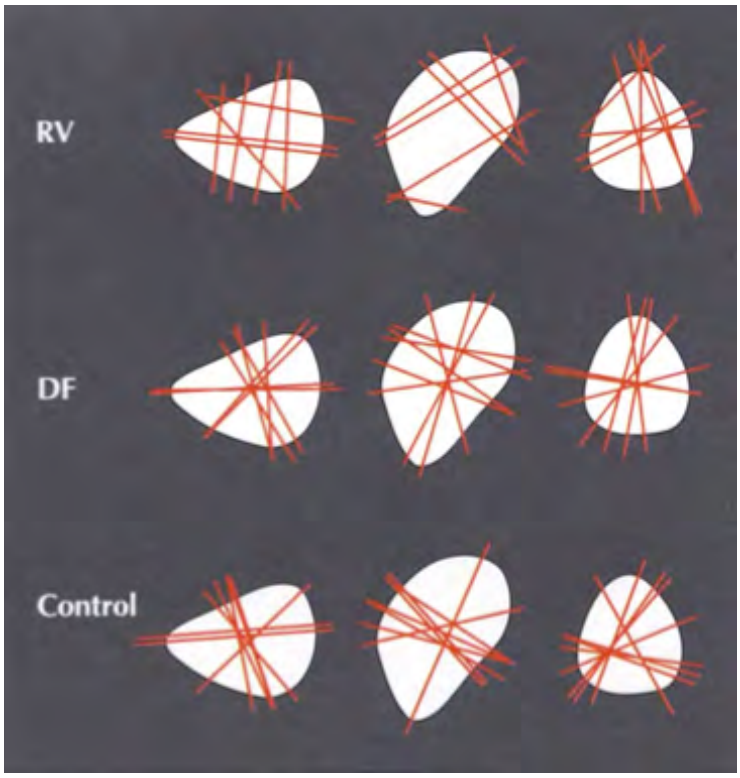

Figure 1. Grasp lines for patients and controls. The "grasp lines" (joining points where the thumb and index finger first made contact with the shape) that were selected by the optic ataxic patient (RV), the visual agnosic patient (DF) and a control subject when picking up three of the twelve shapes.

These converging evidence from early neuropsychological descriptions by Bàlint, more recent ones by Milner and Goodale and the neurophysiological studies in monkeys by Ungerleider and Mishkin have paved the way for the attribution of visuo-motor and spatial abilities in the occipito-parietal cortex.

Turnbull & al. (1997) have proposed a distinction between the spatial and visuo-motor functions of this dorsal (occipito-parietal) visual pathway. They argue that there seems to be an anatomical distinction between the two functions: the superior part of the parietal lobe (or dorsal parietal) where visuo-motor functions would be implemented and the inferior parietal lobule (or lateral parietal) where

spatial information about the scene (objects relative location, distance, size, etc.) would be represented.

The dorsal stream thus seems to critically represent spatial properties of the scene and/or extracts relevant information for motor guidance. Impairments to perceive other properties of object qualities such as color, texture or object category, have been observed in patients and animal studies, for which in areas of the occipito-temporal lobe were lesioned, thus forming another route departing from early visual cortex.

b. Visual agnosia

As in animal lesion studies, which have helped uncover functions of the dorsal pathway, when lesions are done in the occipito-temporal cortex other kinds of impairments arise. Ungerleider and Mishkin (1983) have shown that monkeys were impaired in a short-term visual memory task when they underwent bilateral removal of area TE in inferior temporal cortex. Monkeys were first familiarized with an object and subsequently had to pick out one of two new objects presented which was not the one previously seen. An impairment in this task shows an incapacity to retain and compare visual properties of objects.

Human patients also have been exhibiting this kind of inability to discriminate and compare object properties (Feinberg and Farah, 2000). Some patients were even unable to recognize and/or name objects after certain brain lesions in the occipito-temporal cortex. This disorder has been named visual agnosia, from the Greek *gnōsis* meaning "knowledge" and prefix *a-* meaning "without". The precise definition of visual agnosia is an impairment of object recognition in the absence of intellectual, mnesic or "elementary visual perception" (e.g. brightness and color discrimination, acuity, intact visual fields) deficits. Lissauer (1890) reasoned that visual agnosia can arise in different ways: an incapacity in perceiving the basic features that form objects, thus patients would not be able to see the object as a whole, or an incapacity in accessing the knowledge associated with the perceived object (name, function, etc.).

Apperceptive agnosia

The former deficit characterizes apperceptive agnosia; patients suffering from this type of agnosia can't reproduce basic shapes as in Figure 2. So even though these patients have normal "elementary visual perception", when faced with a recognition or reproduction task where they need to combine the basic visual features of an object presented to them to form a whole shape, they fail to recognize or reproduce it.

Figure 2. Attempts of an apperceptive agnosic patient to reproduce simple shapes. From (Benson D and Greenberg JP, 1969)

Associative agnosia

Another kind of visual agnosia is characterized by the intact ability to reproduce or distinguish different shapes but the inability to associate it with its name, meaning or any other related information, hence the name associative agnosia because of the incapacity to retrieve information associated with a perceived shape. On the other hand these patients are able to recognize an object by its feel in their hands or from a spoken definition, which shows intact general knowledge but impaired access to it from the visual analysis.

Different lesions produce different impairments

Farah (1992) reviewed the different cases of agnosia and compared it with the models or ideas about the object recognition system. Most work assumed a common mechanism for object recognition, as if it was a single type of system that was able to recognize the wide range of objects we encounter. Other researchers emitted the hypothesis of a multi-subsystem organization, each being specialized in a particular category of objects. One of the most extreme proponent of the multi-system view was Jerzi Konorski in his book in 1967 where he hypothesized a 9 domains system: (a) small, manipulable objects; (b) larger, partially manipulable objects; (c) non-manipulable objects; (d) human faces; (e) emotional facial expressions; (f) animated objects; (g) signs; (h) handwriting; (i) positions of limbs, (Srebro, 2012). As we will see in section B) 2) of the introduction this view has been proven quite close to what we currently know of the ventral visual pathway (VVP).

Division of labor

In sum, the ventral part of occipital and temporal cortices seem to be necessary in the perception of object properties which are independent of the conditions in which it is viewed. In other words extracting visual properties such as shape, color, etc. from the visual scene, binding them together, to be able to get a specific representation of an object's physical attributes and access knowledge about it seems to be performed in the VVP. In contrast to the dorsal pathway, which is concerned with the present configuration of a visual scene: where are the different parts of the physical world, how to interact with them, etc.

c. The two-stream hypothesis

The idea of a neural division in visual functions and more specifically a distinction between recognition and spatial, visuo-motor systems dates back to the late seventies by Gerald Schneider (1969) who was studying differences between ablation of a hamster's superior colliculi and visual cortex. In 1982 this hypothesis of two systems was re-explored by Ungerleider and Mishkin using brain lesions in monkeys and they proposed the now famous "what" and "where" streams, corresponding to the recognition and spatial properties of visual cognition

respectively. This view was then updated about a decade later by Milner and Goodale when they reported the case of patient DF, to make the point that the "where" stream should rather be thought of as a visuo-motor stream not subserving spatial localization functions but having a purpose of informing the motor system on how to achieve certain actions using visual information (how to interact physically with an object).
The two-stream hypothesis has been very influential and is still studied to understand exactly how these two-stream do their job and how they interact with each other.

The work presented in this thesis investigated how learning impacts the VVP's functions. Let us now look at functional investigations of neurons in different parts of this visual pathway.

2. Neural implementation

To understand how the neural system underlying visual object recognition is built, let us go through the first steps of light's path into our brain.

a. From light to cortical activity

Visual information is captured by our retinas, a multi-layered lining in each eye and composed, among other types of cells, of photoreceptors. These cells are sensitive to light entering our eyes after being reflected by objects in our environment. They react to this electromagnetic radiation by changing their membrane potential, a signal that is then picked up by ganglion cells that send this signal through the optic nerve to the brain. Different kinds of photoreceptors exist in the retina: cones and rods. Cones are only sensitive to bright (day) light and come in three flavors in the human retina, each sensitive to a different range of the visible light spectrum, giving the possibility to represent the power of different wavelengths of light, i.e. colors. Rods are more sensitive to dim light, which makes them indispensible in low light conditions, i.e. at night.
The spatial distribution of rods and cones is not uniform in the retina, see fig.4. Concentration of cones is highest in the fovea, the central part of the retina, and contains more than half of all cone photoreceptors. Rods are present in the retina's periphery. Because of this organization visual acuity is highest in the fovea

where is projected the center of the visual field. Eye movements compensate this limitation in the periphery by placing the part of the visual field of interest on the fovea.

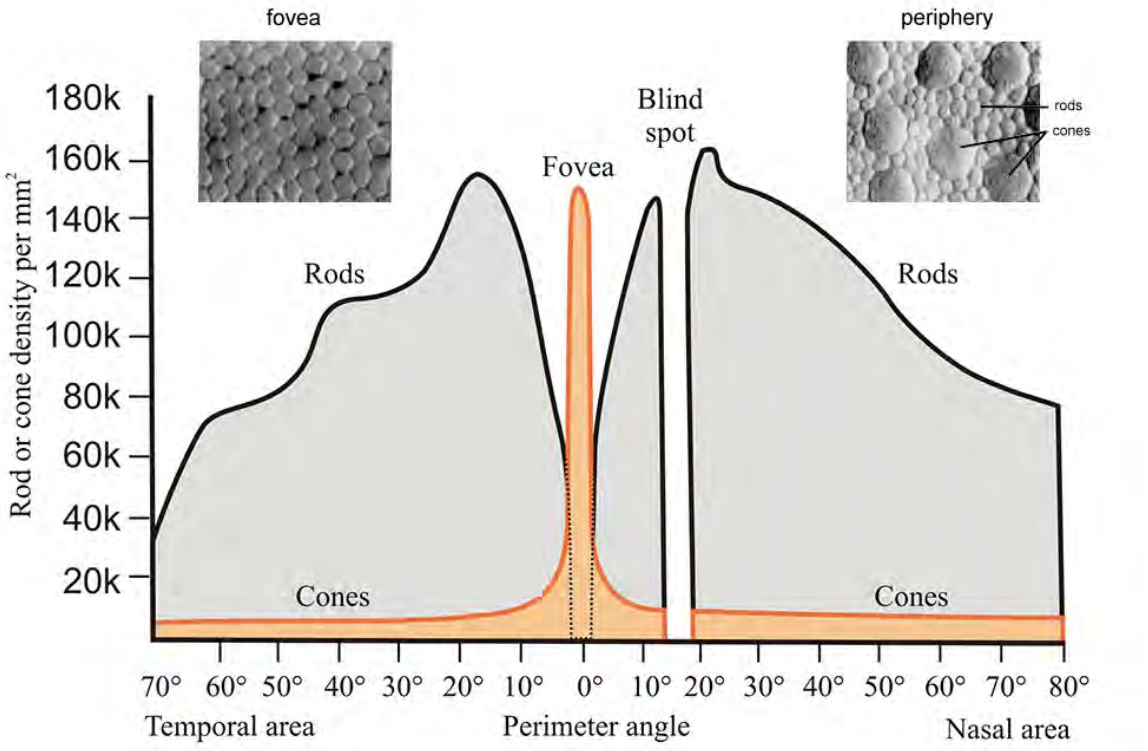

Figure 3. Cones and Rods distribution in the retina.

Photoreceptors constitute the bottom of our three-layered retina. The second layer is composed mainly of bipolar, amacrine and horizontal cells which play different roles in aggregating the photoreceptors' signals. Finally, the last layer of the retina is made of ganglion cells, which receive inputs from cells in the second layer and for some directly from photoreceptors in the first layer. They then send axons that form the optic nerve. After a relay in the thalamus, in the Lateral Geniculate Nucleus (LGN), this information reaches the primary visual cortex (V1). Along this routing of visual information, all cells will share a common functional mechanisms: a sensitivity to specific visual features in a certain part of the visual field, as known as their **receptive field**. These features evolve along the transmission steps from retina to cortex, and through the two cortical stream described earlier, in each cell's receptive field: from the presence of light for photoreceptors to edges and

contours in V1 neurons. This functional characterization will be developed in part B of the Introduction.

Once the information reaches the primary visual cortex, it starts to be routed towards the dorsal and ventral pathway progressively.

b. The V1 fork: the whats and wheres

V1 is one of the best-understood areas of the cortex. It has been studied for more than 150 years and the seminal work by Hubel and Wiesel in the late 50's and 60's has paved the way for the best characterization of a brain area and earned them a Nobel Prize in 1981.

Methods to study brain tissue and V1 specificities

The study of brain tissue can be done in many different ways: anatomically (physical organization at a macro scale like gyri and sulci), physiologically (the study of functions of living systems, e.g. the how and why neurons react to different visual stimuli) or architectonically (anatomy at the cellular level such as types of cells, density of neurons, etc.). All these different investigation techniques helped discover that projections from the LGN land in a topographically defined area of the cortex that has specific architectonical and functional properties. One of them is the *stria* of Gennari, a line (*stria* in greek) visible to the naked eye running parallel to the cortical surface throughout the whole area, hence the **striate** cortex. This particularity comes from the heavy inputs from LGN neurons entering V1 at the fourth of the six-layered organization of the cortex.

Projection to extra-striate areas

Different layers of the cortex subserve different roles in processing and routing neural information. In the primary visual cortex feed-forward inputs from the LGN arrive in layer 4 and feed-forward outputs to other cortical regions depart from layer 2/3. These projections go to different areas of the cortex and as V1 is spatially defined brain cartographers wondered if other similar regions exist in the brain. This cartography of the visual cortex, parts of the brain that responds predominantly or exclusively to visual stimulation, has the goal to define spatially and functionally the different areas that constitute it. As listed earlier there are

different ways to study brain tissue, one of them is by investigating the functional properties: what makes neurons fire.

Retinotopic organization

V1 being the most studied of all visual areas we have learned a lot on its functional organization. For example each hemisphere of the brain encodes half of the visual field, left and right of the fixation: V1 area in the left hemisphere (left V1) analyzes information from the right visual hemifield and right V1 analyzes information from the left visual hemifield.

Another functional organization in V1 is retinotopy. The retina receives a projection of the visual environment onto its photoreceptors; two points close-by in the visual field will stimulate two photoreceptors close-by. This projection is then shipped to the LGN through the optic nerve and subsequently sent to V1 while keeping that spatial relationship. So the retinotopic organization of V1 means that two points close-by in the visual field will stimulate neurons close-by in V1. In other words it is a map of the visual field. It has then been shown that other areas of the visual cortex reproduced this visual field mapping organization.

Increasingly distinct areas along postero-anterior axis between ventral & dorsal streams

Cortical cartographers have used visual field maps to delineate different visual areas; we will see in chapter B of the Introduction that each represents different characteristics of the visual environment.

The closest visual field map to V1 is V2, forming a strip of cortex adjacent and surrounding V1. By studying V2 neurons sensitivity and architectonic properties, Olavarria and Van Essen (1997) have found evidence of a distinction between a ventral and a dorsal part of V2. Such a distinction is reflected more and more as we get further away from V1, forming the pathways defined in part A.2: a dorsal and a ventral route, "what" and "where" streams.

Partitioning visual areas is a complicated endeavor for many reasons: the boundaries between areas can be very subtle, e.g. where does the cell type significantly changes between striate and extra-striate cortex, internal heterogeneity, e.g. gradual shift in functional properties of an area's, or landmark

variability, e.g. inter-individual differences in sulci and gyri morphologies. But even though there are very different partitioning schemes there are many similarities between the delineation that are yielded suggesting a certain degree of correspondence between some structural and functional measures (Van Essen, 2004).

For many technical and theoretical reasons the most used partitioning scheme to study the visual system by area segmentation in human is visual field mapping. It has led to the discovery of many maps on both brain hemispheres, each representing one hemifield. In the following section we will review the organization of these areas in the ventral pathway.

c. A constellation of areas: partitioning by visual field representations

As we discussed previously, many methods exist to delineate visual areas: cytoarchitecture, connectivity, functional properties, etc. But most of these criteria are unavailable in vivo for the human brain or not well understood. Visual field maps in the cortex on the other hand is a simple and straight forward concept and can be measured relatively quickly and easily using neuroimaging techniques such as functional Magnetic Resonance Imaging (fMRI). This technique allows us to record indirectly brain activity by measuring oxygen consumption in different parts of the brain with millimeter-scale resolution, see chapter 2 for details on this technique.

One popular paradigm is the traveling wave method. Subjects are instructed to fixate a cross at the center of a screen while contrast patterns are shown. These stimuli appear at different eccentricities and angles, covering much of the visual field (from 8 to 20° of visual angle depending on the study). By recording their brain activity using fMRI, it is then possible to build maps of cortical sensitivity to spatial location in the visual field. By definition each visual field map contains a single cortical location sensitive to a point in space, so two cortical locations sensitive to the same point in space must belong to different visual field maps, i.e. different visual areas.

Using visual field mapping multiple areas have been described. The first cluster of maps contains V1, V2 and V3. These three areas have been delineated and studied well before the invention of the fMRI retinotopy (Zeki, 1969), which makes them a good control for the validity of this technique. Areas V2 and V3 in each hemisphere represent the entire contralateral visual hemifield, although each quadrant of the hemifield, i.e. upper and lower, is represented on the ventral and dorsal part of these areas respectively, see Figure 4. This organization led to the definition of subdivisions between an occipito-parietal part, e.g. V2 dorsal (V2d), and an occipito-temporal part, e.g. V2 ventral (V2v), same thing with V3. These areas are often referred as "early visual areas" or "early visual cortex" (EVC).
Beyond the EVC other visual field maps can be grouped into occipito-parietal and occipito-temporal areas, being part of the dorsal and ventral stream respectively.

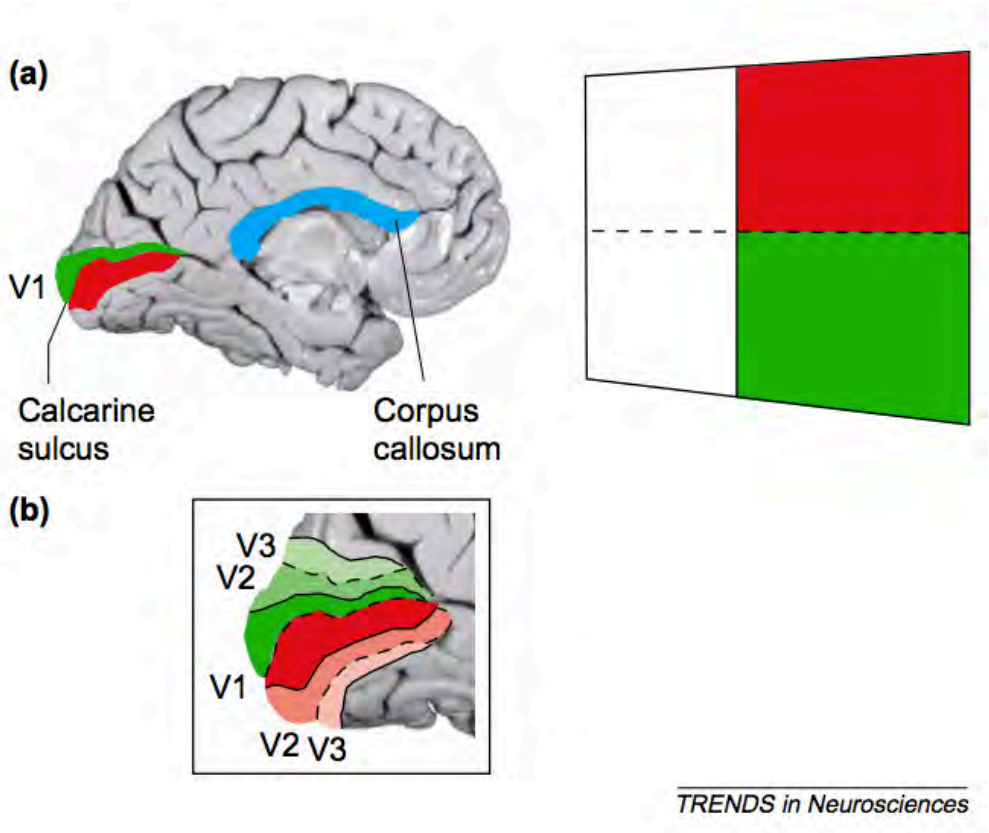

Figure 4. Early visual areas and visual field coverage. (a) Represents the parts of V1 which will be stimulated by each quadrant. (b) The quadrant represented by areas along the ventral and dorsal streams is represented by different greens and reds for lower and upper field quadrant of the contralateral hemifield. From (Zeki, 2003)

The occipito-temporal visual field maps are usually grouped into lateral occipital (LO) and ventral occipital (VO) maps. The visual field coverage varies depending on the areas. Throughout the VVP areas respond more to stimuli in the central part of the visual field. There is an increasing cortical magnification of the center of the visual field, meaning that inputs from the fovea are magnified, they occupy a greater cortical surface, compared to peripheral inputs (Baizer et al., 1991). It has been shown in monkeys and humans that increasing the stimulus radius from 3 to 16 degrees will expand the responding surface in V1 considerably but ventral occipital maps expand very little.

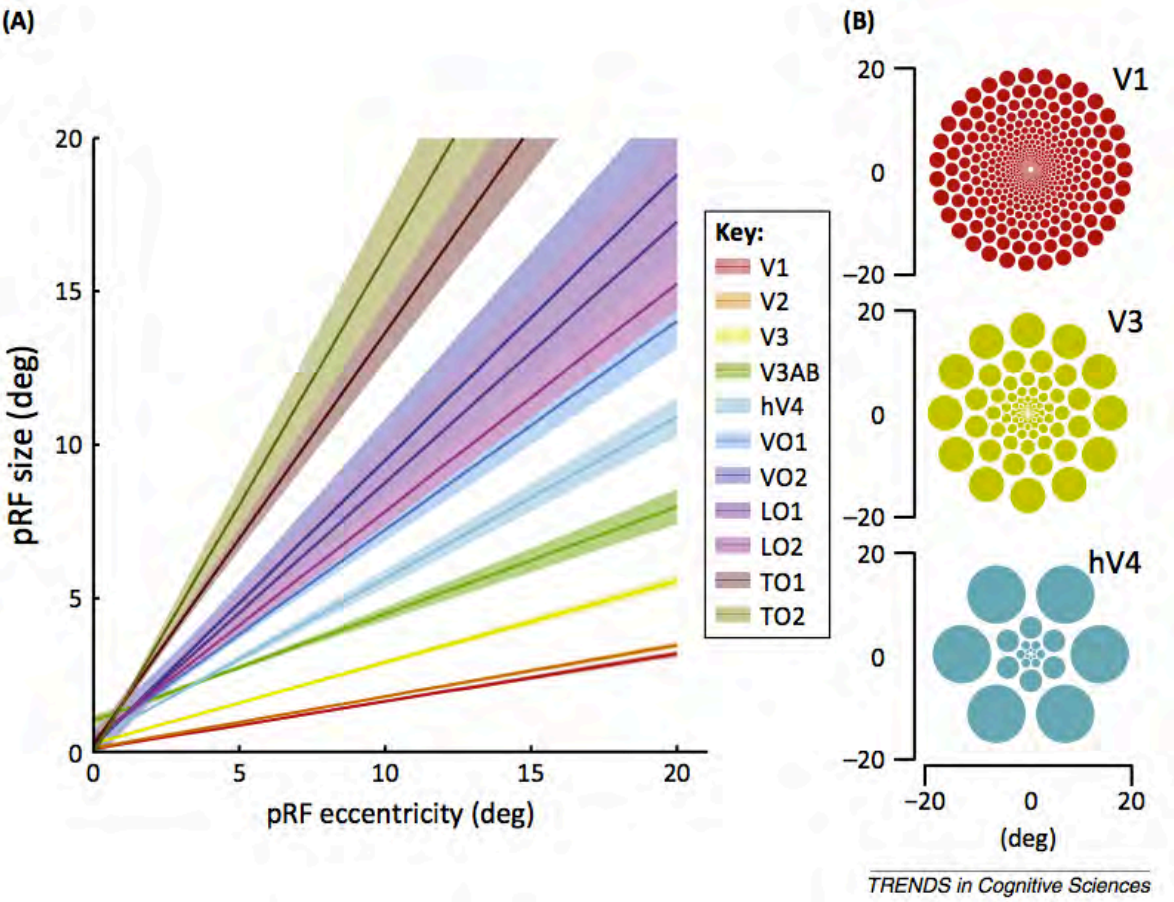

Figure 5. RF size along the ventral pathway

Receptive fields have been studied across the entire visual cortex for a long time and it is now possible to estimate the RF of voxels belonging to each area using fMRI (Wandell and Winawer, 2015). Because voxels contain thousands of neurons this approach is named population receptive fields (pRF). As with previous

electrophysiological studies it has been shown that within an area, receptive fields sizes vary: from foveal to peripheral they get bigger, i.e. a neuron sensitive to visual field location at the periphery will have a pRF larger than one sensitive to a foveal location. This increase in pRF sizes is observed along the VVP, some areas having almost no retinotopic organization and being sensitive to complex features present in the visual field.

Even in EVC stimuli inside a neuron's RF do not always elicit an increase in spiking activity. It has to be the right stimulus, the feature that this neuron is sensitive or tuned to, and these features increase in complexity along the VVP.
Together the increase in receptive field size and complexity of feature sensitivity led to the idea of a hierarchical organization in the visual system: going from V1 to infero-temporal cortex.

B. The ventral stream: hierarchical organization feeding the MTL

The original notion of hierarchical processing in the visual cortex was put forward by Hubel and Wiesel while they were studying V1 (Hubel and Wiesel, 1962). While describing physiological properties of cells in different layers of V1 they discovered an increase in complexity in their receptive fields. The hierarchical organization of visual processing has since then been tested using many different techniques (Felleman and Essen, 1991) and even though its strict formulation as a uniquely feed forward operation is impossible, because of the enormous amount of feedback connections at different levels of the hierarchy, it is still a well recognized framework for the study of the visual system and has inspired most computational models of vision (Serre et al., 2007; VanRullen et al., 2001).

The hierarchical processing is thought to progress along the IT cortex that encounters the medial temporal lobes (MTL) which have been shown to support learning and memory functions, which we will describe in part **C**. of this chapter.

Understanding the representations of visual stimuli at different levels of the hierarchy has been a challenge of the past decades. The knowledge on the spatial arrangement of functional areas, the functional modules it is composed of and the temporal dynamics of this process has progressed tremendously. Thanks to advances in experimental design, neuroimaging techniques and analysis tools it has been possible to extract comprehensive information from the colossal amount of data collected since researchers have started trying to understand the VVP.

In this chapter we will go along the VVP's hierarchy to explore its representations and the techniques that helped to understand object visual recognition in the brain.

1. From oriented bars to shapes

The central point of Hubel and Wiesel initial finding (Hubel and Wiesel, 1959) was that oriented slits of light were the preferred form of visual stimuli for eliciting activity in V1 neurons. That is each neuron was tuned to a particular orientation of a line of light in their RF. What also made their discovery incredible is the fact that

they related it to the LGN neurons sensitivity to spots of light by elaborating on the integration of LGN neurons' RF into V1 neurons' RF.

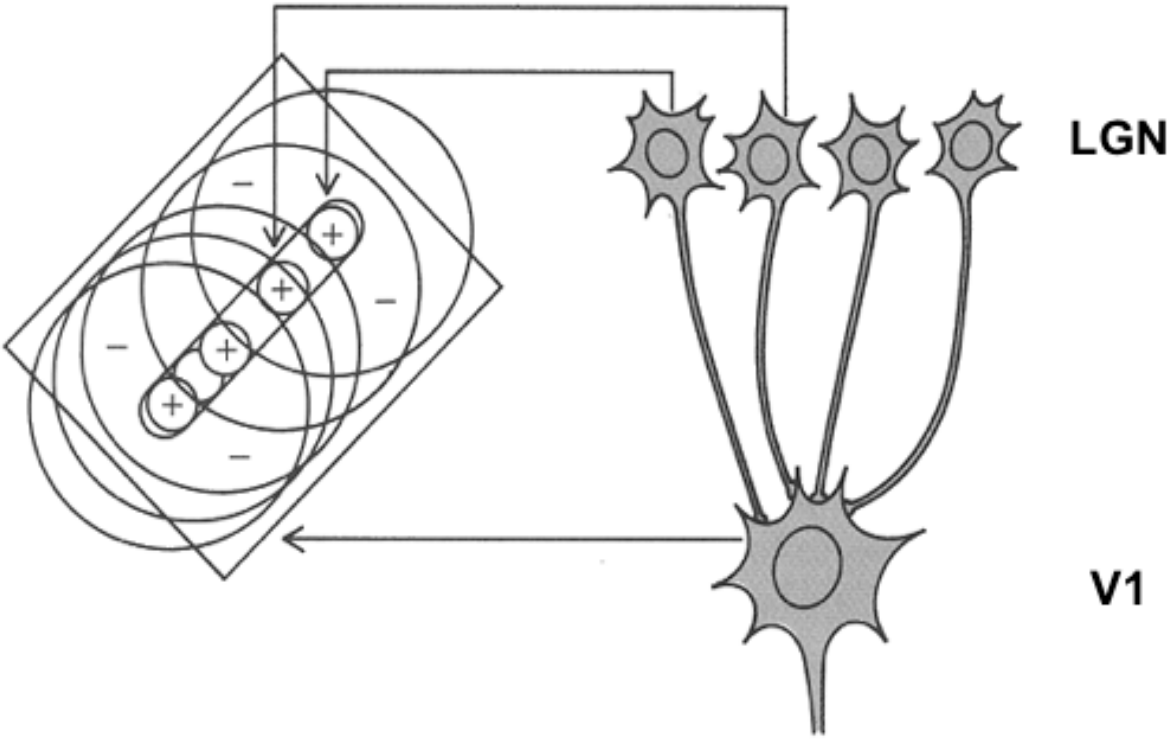

Figure 6. Receptive fields integration from LGN to V1.

While recording from neurons in different layers of V1 they were able to distinguish two types of cells: simple cells, as we just described, and complex cells, which had a similar preference for orientation but were invariant in the position, or phase, of the line inside their receptive field. The integration of LGN RFs into single V1 neurons and this increase in invariance from simple to complex cells in V1 was one of the key features that led them to posit a hierarchical organization in the visual system.

One of the earliest and most prevalent models of V1 simple cell receptive field is the one proposed by Marčelja (1980) and tested by Jones and Palmer (1987). They described V1 simple cells receptive field as a specific kind of linear filter, i.e. a Gabor filter.

A Gabor filter is the product of a two-dimensional Gaussian, defined by specific mean and variance, with a sinusoidal grating with specific frequency, phase and

orientation. In practice the frequency and size are tied. Figure 8 depicts a range of Gabor patches that vary in orientations on the x-axis, and size on the y-axis. The increase in invariance from simple to complex cells is easily represented using the Gabor model by spatially shifting the sinusoid while keeping the Gaussian still.

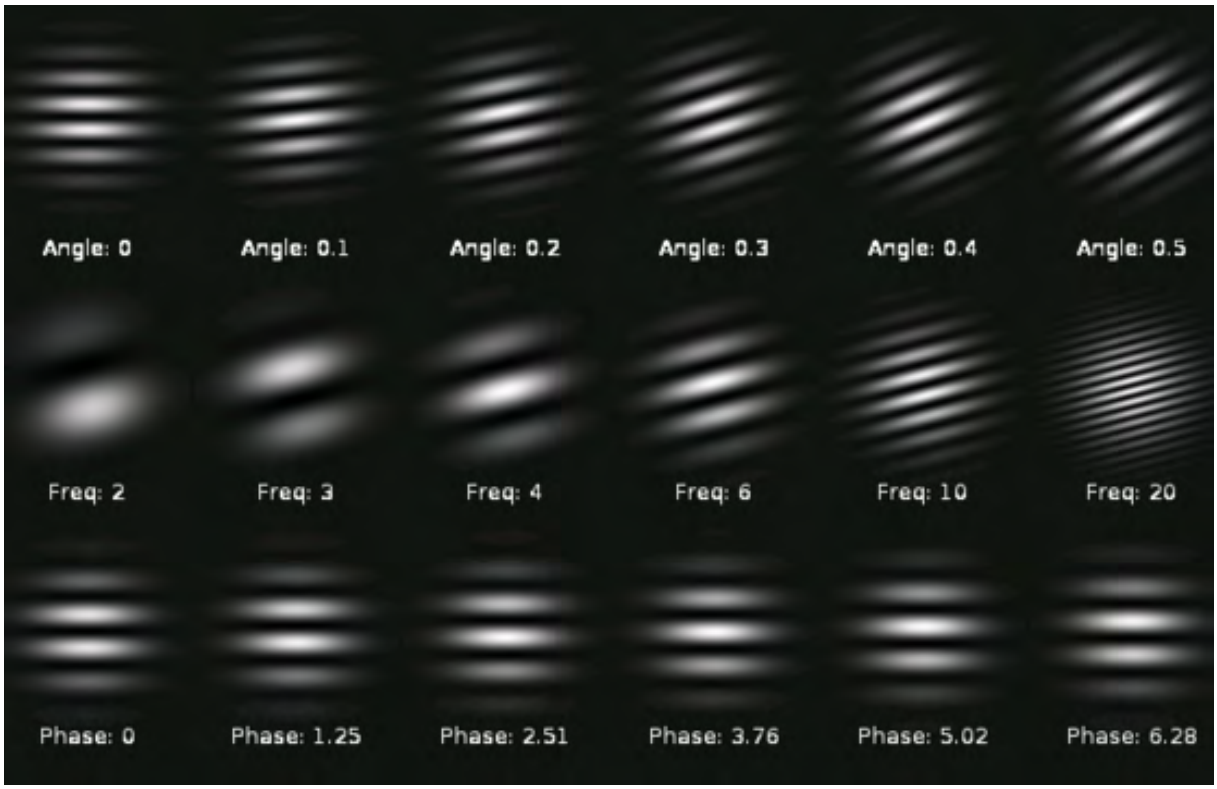

Figure 7. Gabor patches. Variations in orientation, spatial frequency and phase of gabor patches.

More recently, using elaborate machine learning methods, a dictionary of Gabor patches and fMRI recordings, researchers have been able to fit the Gabor computational model of simple and complex cells RF to each voxel in EVC (Kay et al., 2008). This work has shown that quantitative evaluation of voxels in EVC is possible and that this model catches a great amount of fMRI activity variance, but decreasing from V1 to V3, indicating that features represented in later stages of the EVC were not as simple as those represented in V1.

Even though we will focus on selectivity to form in this overview of the literature, V1 organization is far more complex than a visual field map with cells sensitive to

oriented bars in the visual field. Among other things, V1 represents colors, ocular dominance input (which eye is predominantly projecting to a particular population of V1 neurons) and motion direction.

In later stages of the hierarchy it has been more challenging to characterize tuning dimensions of visual information. Kourtzi and Connor (2011) have hypothesized that the difficulty in understanding neuronal tuning in the VVP has come from the fact that the dimensionality of the object domain is too high and that there is no obvious, straightforward way to represent complex object into neural response. What has been done is to sample the stimulus space by sampling arbitrary features of natural scenes like objects or simpler ones like contours or shapes to evaluate sensitivity of these stages of the ventral stream.

V4 sensitivity to components of shapes

As Kay & al. (2008) have observed and as has been shown before (David et al., 2006) the use of Gabor models does not predict accurately the activity elicited in intermediate areas such as V3 or V4. Using non-linear models of orientation and spatial frequency spectrum David & al. (2006) showed that compared to V1, V4 neurons were selective for non-Cartesian gratings, i.e. not a simple right angle arrangement of straight lines but spirals and concentric curves. They also showed evidence of bi-modal orientation tuning for some V4 neurons, meaning that across the area some neurons are selective to conjunction of orientations of different angles and smoothness (i.e. sharp or rounded angles), forming contours. Pasupathy and Connors (2002) have shown that V4 neurons encode shapes in terms of their constituent boundary features by representing different kinds of contours in an object-centered manner. This result supports the notion of shape coding by parts-level (or components) representation in V4.

Later areas sensitive to complex objects

Beyond V4 receptive fields become larger and their tuning properties even harder to characterize. But this part of the cortex, also known as the higher-level visual cortex because of its hypothetic level in the visual hierarchical processing, is sensitive to more perceptually meaningful stimuli, such as object categories, views of different objects, etc.

2. Object category selectivity

Taking the visual recognition problem from the other end has proven very fruitful to study higher-level visual cortex, that is: using natural stimuli instead of using artificial low-level visual features. Starting in the 70's, studies investigating selectivity of neurons in IT found that some neurons responded to faces and other objects (Gross et al., 1972). Later on, using neuroimaging techniques, it has been possible to characterize brain regions that respond to perceptually clustered stimuli such classes of objects (e.g. cars, letters, chairs or faces). Multiple patches of category selectivity have been established using fMRI: for faces in the Fusiform Face Area (FFA) (Kanwisher et al., 1997), places and scenes in the Parahippocampal Place Area (PPA) (Epstein and Kanwisher, 1998), visual forms of words in the Visual Word Form Area (VWFA) (Cohen et al., 2000) or an area selective to body parts in the Extrastriate Body Area (EBA) (Downing et al., 2001). These areas, which can be overlapping depending on protocol design or analysis techniques, have been the focus of a great debate concerning their organization of IT in modules or in feature maps with regions representing features useful for detecting particular stimuli classes (see Op de Beeck et al. (2008) for a review).

a. Object selectivity at the neuronal level

The seminal study by Gross et al. (1972) and later work (Desimone et al., 1984) has found evidence of neuronal sensitivity to complex stimuli such as faces and objects.

Later on Keiji Tanaka's thorough work on object selectivity in IT (Kobatake and Tanaka, 1994; Tanaka, 1996) unraveled the organization of these selectivity in columns of visually similar stimuli such as different views of an object. The viewpoint selectivity of neurons in IT decreases along the VVP. As for the tolerance to oriented bars position in the visual field in V1 from simple to complex cells, there is a increasing tolerance, also termed invariance, to some stimulus dimensions such as viewpoint, location in the visual field, etc.

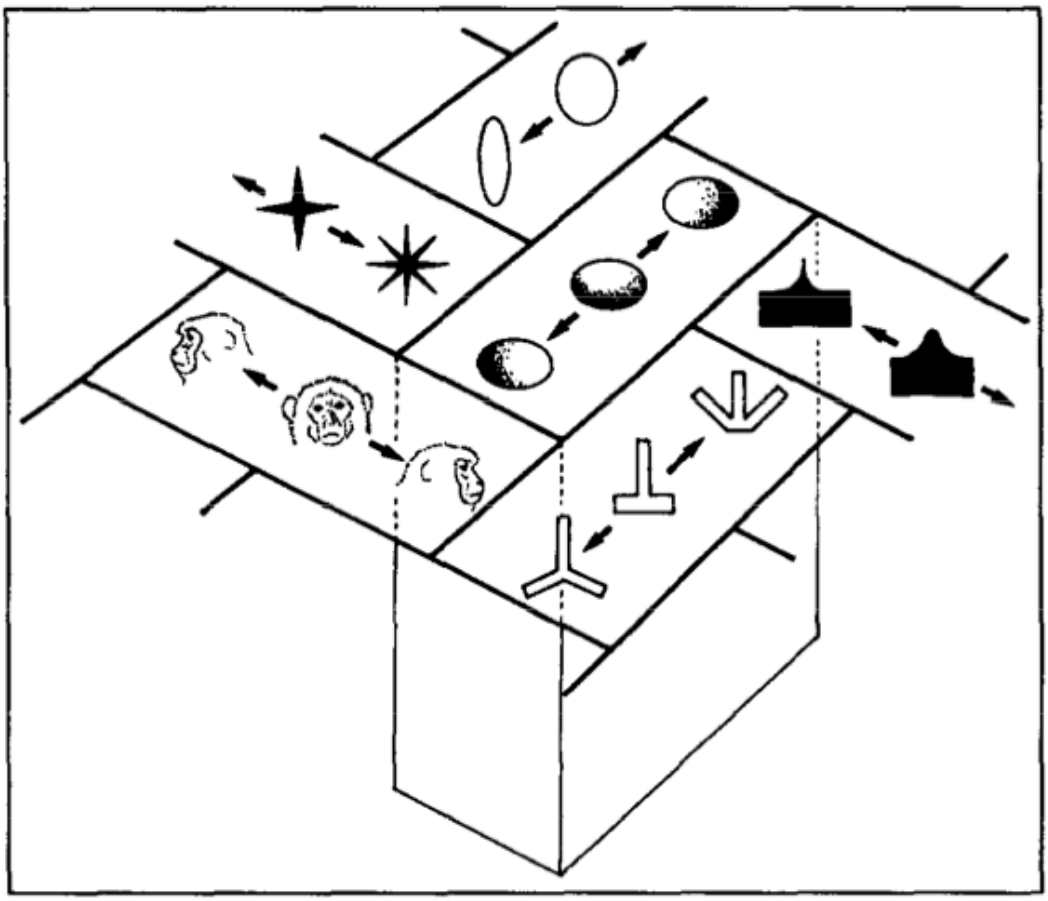

Figure 8. Schematic diagram of columnar organization in IT. Columns representing different features such as families of shapes (circles, spheres, conjunctions of lines, etc.) or views of a certain category of stimuli. From (Tanaka, 1996)

These neurons having great invariance and selectivity to very complex features can be found in the anterior temporal lobe, near the MTL, a region critical for memory formation (Squire and Zola-Morgan, 1991). Miyashita and colleagues (Miyashita, 1988; Sakai and Miyashita, 1991) have tested the plasticity of these neurons' selectivity in a series of associative learning experiments. Their results indicated that they could become selective to pairs of visual stimuli through long-term associative learning between them and that neurons, which were selective to one of the paired stimuli from the beginning showed an increase during the delay period when the other stimuli of the pair was presented as a cue. These findings indicate that neurons at these stages of the ventral stream hierarchy could tune their selectivity to newly learned stimuli or associations.

b. Spatial layout of object selective cortex: areas of preferred categorical selectivity

Single-cell studies in non-human primates have uncovered object selectivity and columnar organization in the temporal cortex, i.e. VVP (Gross et al., 1972; Kobatake and Tanaka, 1994). These studies demonstrated selectivity to whole, real world objects (e.g. faces, man made objects) compared to selectivity of earlier stages of the hierarchy (e.g. EVC, V4) to "basic" visual features (e.g. oriented bars, curvature) which represents a major step towards understanding of the neural implementation of visual perception because it relates neural activity to perceptually meaningful stimuli.

Nevertheless it does not provide a description of the large-scale organization of the ventral visual pathway because of electrophysiological limitations in sampling.

Thanks to the advent of functional MRI an account of the large-scale organization of the VVP has been possible. fMRI has revealed a topographical organization for specialization as demonstrated by localized increases in neural activity, i.e. levels of blood oxygenation in fMRI, to classes of stimuli such as faces, places, limbs and other objects.

Large-scale study of IT selectivity using fMRI

In the mid and late 90's the advent of fMRI gave researchers the possibility to unravel large-scale brain responses underlying object recognition. Malach & al. (1995) compared activation induced by pictures of objects and a wide range of texture patterns and found preferential activation for the former category of stimuli in lateral-posterior occipital cortex. This area, termed Lateral Occipital Complex (LO), was sensitive to the degree of object visibility but not other low-level properties such as visual size, suggesting a role of LOC in representing objects in contrast with low-level properties of the stimuli. In this study they defined LO as the voxels which exhibited a significant preferential activation to pictures of objects compared to textures. This procedure is called a **localizer**. A few years later Grill-Spector & al. (1998) provided a more detailed investigation of LO functional properties, comparing its response to earlier stages of the visual hierarchy (e.g. V1, V4v, V3A) and showed that by scrambling blocs of pixels of natural images the

response of visual area decreased from V1 to LO as the structure of the image was deconstructed, i.e. degrading the object information content. By averaging the signal across all the voxels contained in an area, e.g. V1 or LO, they were able to compare activity levels in a **univariate manner** between conditions.

This localization of a patch of cortex specialized in processing objects using fMRI launched a series of studies which discovered several other cluster of categorical specificity.

Using localizers and univariate analyses Kanwisher's group was able to define three new category specific regions selective for faces (FFA), spatial layouts such as scenes or places (PPA) and body parts (EBA) (Downing et al., 2001; Epstein and Kanwisher, 1998; Kanwisher et al., 1997).

Not long after the "discovery" of the FFA, other groups have argued that this area was not only category specific. Gauthier & al. (1999) designed a study where they tested the hypothesis that the FFA was an area recruited by expertise on any type of objects and that the face specificity of the FFA was due to our particular expertise with this kind of stimuli. Subjects were trained at categorizing novel objects, i.e. greebles, until reaching certain criteria considered to be diagnostic of expertise. Acquisition of expertise for greebles led to increased activity in the right hemisphere face areas for experts compared to novices. This result was replicated in cars and birds experts showing that the degree of expertise towards an object category predicted relative activation of the right FFA, strengthening their argument that the level of expertise was more determinant than superficial object properties in the FFA (Gauthier et al., 2000).

One thing to notice is that all these studies used a similar analysis method which is localizing voxels clusters which respond significantly more to a particular category of objects, i.e. a localizer, and compare the average response across all these voxels to different kinds of stimuli categories or tasks. We will see in the next part (I.B.2.c.) that using the multivariate nature of neuroimaging data, e.g. fMRI, permits to unravel other information compared to the analysis of univariate signals, i.e. averaging voxels response across an area.

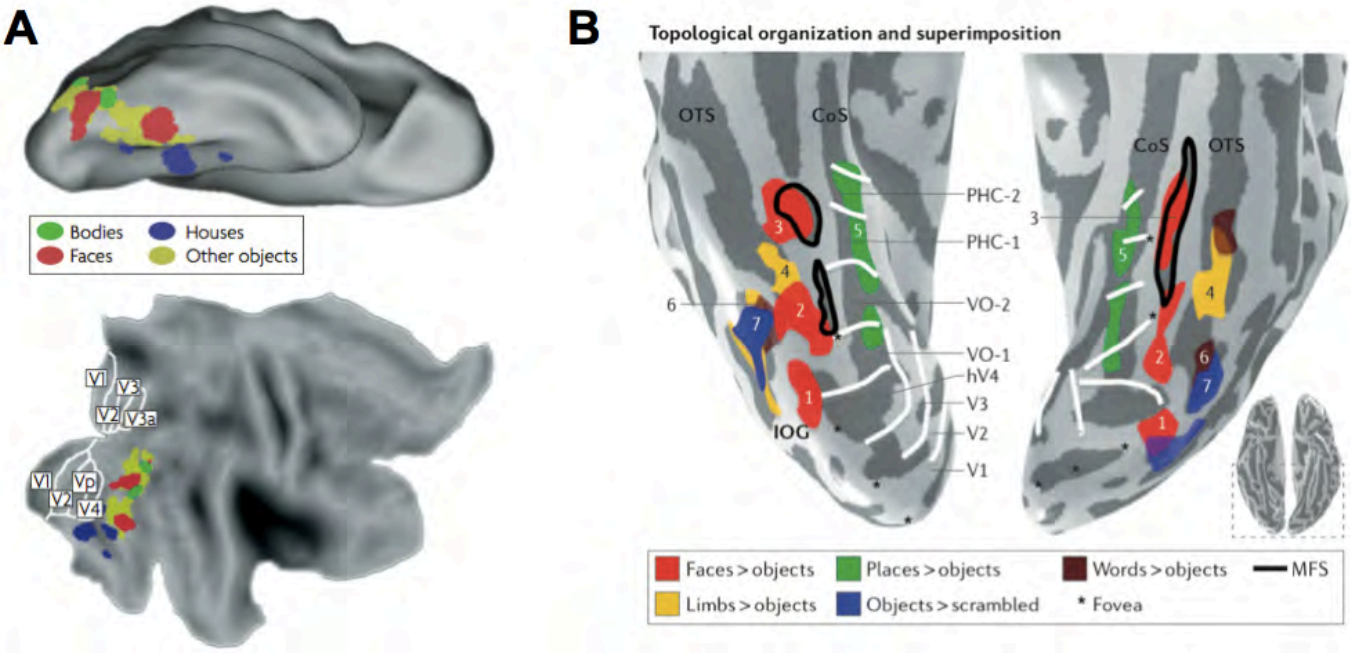

Figure 9. Typical location of retinotopic and category selective regions. (A) Typical locations of category selective regions in the human ventral visual cortex (left hemisphere). Top schema depicts object category preference areas on an inflated view of the left hemisphere. Bottom schema depicts a flattened view of the left hemisphere with object category preference as colored areas and outline of early visual areas. From (Beeck et al., 2008). (B) Depicts similar areas on an inflated view of both hemispheres with the addition of anatomical landmarks and retinotopic maps. OTS: Occipito-Temporal Sulcus; CoS: Collateral Sulcus; IOG: Inferior Occipital Gyrus; MFS: Mid-Fusiform Sulcus; PHC: ParaHippocampal Cortex; VO: Ventral Occipital. From (Grill-Spector and Weiner, 2014).

The view of the VVP as a modular organization whether it processes specific stimuli categories or subserves particular perceptual abilities, e.g. expertise, is an ongoing debate. Models for this functional architecture have been formulated and can roughly fall into three families:

- **The module family**: The object selective cortex is composed of modules, each allowing recognition of a particular class of stimuli, e.g. FFA for faces.

- **The process family**: Different regions underlie perceptual processes. In particular this view considers the FFA as an expertise area in discrimination of individual exemplars (Tarr and Gauthier, 2000).

- **The distributed family**: Representations of object categories is overlapping and distributed over the entire ventral temporal cortex. This representation would be based on attributes of object forms that compose different classes of objects.

c. Multivariate analysis of object representations in the VVP

Haxby and colleagues (2001) realized one of the first study testing the third model. Instead of contrasting conditions, e.g. display of faces versus houses, they measured the correlation of fMRI activity patterns between and within conditions. Even when excluding voxels that fell into category specific areas they were able to show that categorical information was not only present in the area that responded the most to this class of stimuli but distributed across IT. Their work was one of the first that made a historical shift between univariate and multivariate analysis of functional imaging data, i.e. using one variable to represent the information content of an area versus using the multivariate pattern of activity (MVPA) respectively. We will further develop the topic of MVPA techniques in section B.3 of the introduction.

In this section we have reviewed the large-scale spatial organization of category selectivity in the VVP using univariate analysis of fMRI data, see Figure 9 for summary of the areas on inflated and flattened surface of the VVP. This technique gives largely independent information about the activity in different areas of the VVP but the indirect neural response it captures through the blood-oxygen-level dependent (BOLD) signal has a temporally sluggish resolution (see section III.A.1 for more details). Therefore it gives us little information on the dynamics of perceptual processing in the VVP, which can take place in tens or hundreds of milliseconds.

d. Temporal dynamics of the visual system

As we have seen in section A.1 of the introduction, the VVP serves visual recognition but another constraint weighs on these processes: time. In fact a visual

system which recognition abilities are "perfect" but needs several minutes to achieve this operation would not be ecologically relevant, for example for detecting a predator in the scene. These temporal constraints are also interesting to fully understand and characterize the computations realized by the VVP.

Category selectivity in scalp topographies and ERPs

Even though non-invasive electro-physiological recording techniques, i.e. EEG and magnetoencephalography (MEG), have a poor spatial resolution compared to fMRI, their high temporal resolution, i.e. at a millisecond scale, allow to observe strong category selective neural responses in the time-course and scalp distribution of the recorded neural activity.

The largest and most consistent event-related potential (ERP) difference between object categories has been observed as early as 130-170ms at occipito-temporal recording sites, for example an increased negative peak at 170ms is consistently present for face stimuli compared to objects, i.e. the N170. The topography of ERPs also reflects differential processing of visual categories, for example a hemispheric advantage for the processing of faces in the right hemisphere and words in the left hemisphere which is consistent with fMRI studies which discovered laterality differences in category specific regions of the VVP (Cohen et al., 2000; Kanwisher et al., 1997).

Neuronal latencies of areas in the VVP have shown progressive temporal responses from EVC in a sequential way until later stages of processing in IT, e.g. TE in monkeys, which roughly corresponds to IT. There are discrepancies in the strict hierarchical view of the visual system, e.g. latencies of cells in V2 can vary depending on the recording site, for some earlier than some types of cells in V1, in part due to different neuronal pathways from retina to visual cortex (Nowak and Bullier, 1997). Nevertheless many studies have shown that a first feed-forward sweep of visual information can account for rapid categorization. One such example was a seminal study by Thorpe & al. (1996) who were able to estimate the processing time necessary for fast categorization of animals in briefly flashed natural photographs. Using electroencephalography (EEG) they were able to show large differential activity in the evoked EEG activity between targets and distractors as early as 150ms after stimulus onset.

Following these results another study has shown that these very early categorization abilities were not only possible for evolutionary relevant stimuli such as animals or landscapes. These types of stimuli could be special cases for the visual system and might rely on hardwired innate mechanisms. VanRullen & Thorpe (2001) showed that even for stimuli having less biological significance and being more recent in our environment such as means of transports like cars or airplanes, ultra-rapid visual categorization was possible and undifferentiated in terms of accuracy or reaction times. This result also shows the learning capacities of the visual system, which can be molded to recognize artificial stimuli as well as natural ones based on experience.

One tremendous advantage of neuroimaging methods having high temporal resolution is that allows us to study different rhythms of neural activity in the brain. Different rhythms, i.e. frequencies, of neural activity reflect different neural processes or arousal states (Buzsaki, 2006). For example deep stages of sleep are characterized by slow frequencies below 4Hz, i.e. delta band, while conscious perception of masked visual stimuli is associated with activity in the gamma band between 35 and 60Hz (Summerfield et al., 2002).

Vidal & al. (2010) investigated the visual object category selectivity of different rhythms: high frequency gamma-band activity (50-150hz), low-frequency alpha/beta-band activity (8-24hz), and the more traditional ERPs using intracerebral recordings in epileptic patients. They showed that both gamma-band activity and ERPs exhibited strong category selectivity and that there was very little spatial overlap between sites eliciting the same category-specificity. Even though they don't elaborate much on the mechanisms underlying the spread of category selectivity among these neural markers, it shows that studying the information content of different neural activity rhythms can lead us to better understand the neural code underlying visual recognition, which we investigated in chapter II.

3. Univariate vs Multivariate

Univariate analysis of fMRI and electrophysiological data has helped understand the spatial and temporal organization of object selectivity in the VVP. The basic method was to compare the average activity between conditions, e.g. activity elicited by faces versus houses or objects versus scrambled images. New technological advancements in analysis of neuroimaging data have transformed the approach from measuring activity to information by using the multivariate nature of the signal, e.g. from averaging the activity in the FFA to comparing voxel patterns of activity in the area between conditions.

a. Classifying patterns of activity

Classification techniques represent a sub-group of supervised machine learning algorithms. The basic idea behind it is an algorithm that will try to find the discriminative features between two classes of patterns, e.g. patterns of brain activity. These algorithms will go through two phases:

- **Training**: the classifier is fed m data points in n-dimensions, i.e. patterns, each accompanied by a label, i.e. class A or class B. The classifier then uses optimization methods to find how to separate the points belonging to each class, e.g. find, in a 2-dimensional space, the line that best separates class A points to class B points.

- **Testing**: the classifier is fed new data points and, based on the criteria found in the learning phase, classify them, e.g. into class A or B. The performance of the classifier is how many times it got the classification right.

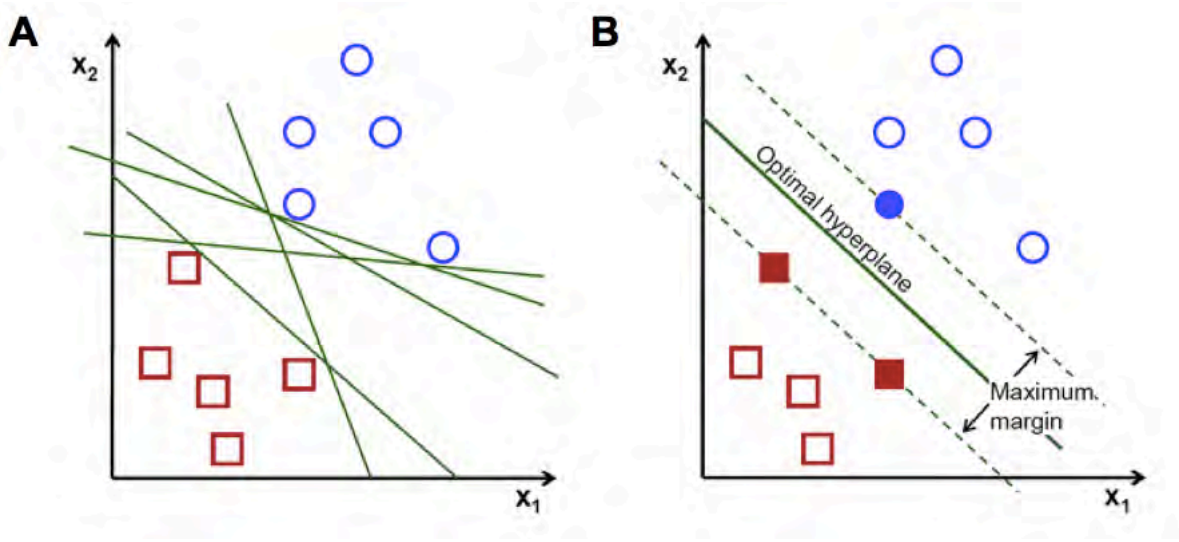

Figure 10. Example of a Support Vector Machine (SVM) classifier on a 2D dataset. (A) The red squares represent one class of points and the blue circles the other class. The green lines represent all the possible delineation the classifier could use to separate the 2 classes. Different flavors of classifiers will use different strategies to decide which line (in a 2D dataset) or hyperplane in a n-dimensional dataset) to use. (B) One particular type of classifiers are Support Vector Machines (SVM) which attempt to find the hyperplane which will maximize the distance to the closest training points. (Adapted from http://docs.opencv.org/).

b. Using machine learning techniques to study visual representations

As we evoked in section B.2.b traditionally neuroimaging data analysis was done by trying to find significant responses to an experimental condition (or differences between conditions) by focusing on measurement channels one-by-one, e.g. electrodes in EEG studies or voxels in fMRI studies. This univariate method then averaged across the measurement channels which exhibited the same type of responses to increase the signal-to-noise ratio (SNR). Even though this approach has been tremendously productive it has some limits on what can be tested about the neural code. While this method reduces the noise it also reduces signal from weaker, and thus non-significant, channels which still carry some information about the experimental condition. It blurs the spatial structure of the neural responses which might contain subtle information.

In the early 2000's a major realization has been brought by Haxby et al., (2001). They were among the first to use the multivariate nature of the data by comparing patterns of activation instead of mean level of activation in an area. Using this

approach they were able to classify which pattern of activity was evoked by which stimuli category. Other methods exist to compare multivariate data such as machine learning techniques. Their way of analyzing the data thus opened an enormous amount of new tools available in the machine learning literature to study the presence of a specific information or representation, e.g. difference between visual categories and how it is structured. Contrary to univariate methods, MVPA uses the small and large sensitivities of each channel and in some cases, using machine learning techniques such as SVM, is able to weight the different channels to extract the most information from each pattern of activity.

During the past 15 years MVPA has been used to study covertly attended stimuli information (Kamitani and Tong, 2005), covert episodic memory recall (Chadwick et al., 2010) or conscious perception (Haynes and Rees, 2005). Critically MVPA might be able to tell the difference between activity patterns elicited by two conditions even if the average level of activity does not differ between them.

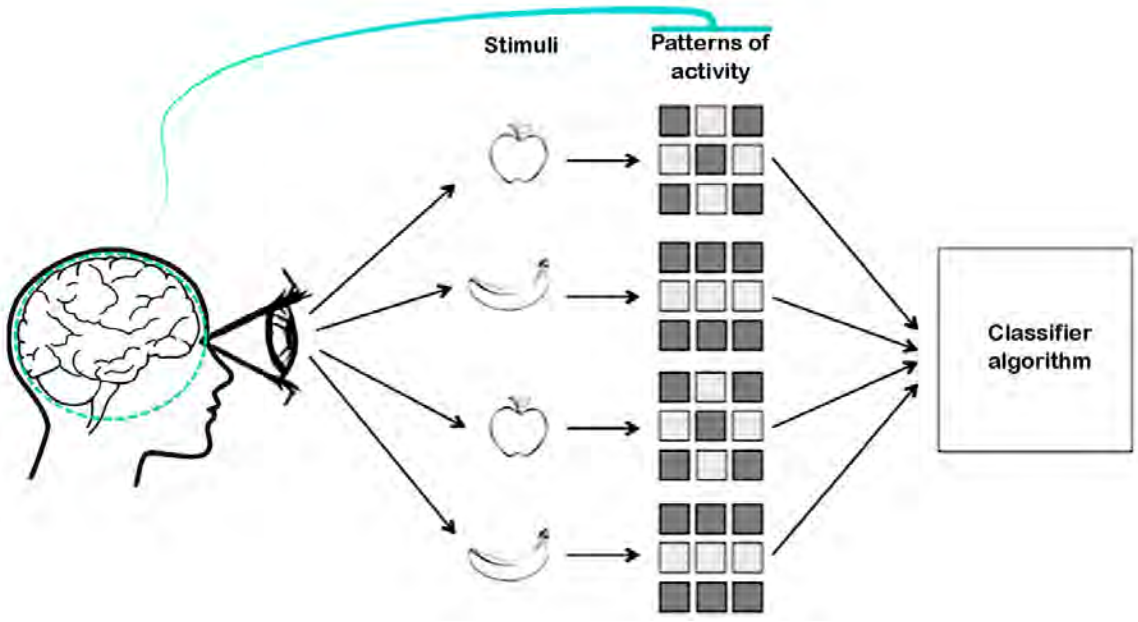

Figure 11. Inputs to classification algorithms. Example from a visual category neuroimaging experiment. Adapted from (Kaplan et al., 2015)

MVPA better than Univariate!

Harrison and Tong (2009) used MVPA to explore the engagement of EVC in working memory (WM). Subjects were shown two gratings sequentially, then a cue

to indicate which one should be kept in WM. They found that the EVC participates in WM, and more surprisingly that the average bold decreased over time but the information content revealed by MVPA was sustained throughout the delay period showing a sensitivity advantage of multi-variate methods.

“Hi-jack” a classifier to study pattern similarity across conditions

As we emphasized above one of the main goal of cognitive neuroscience is to investigate the nature and structure of neural representations. Thanks to MVPA it is possible to test if two visual processes share similar neural representations in the VVP across cognitive states by using classifiers to probe. In an fMRI study Reddy et al., (2010) tested the hypothesis that perception and imagery of visual categories shared similar neural representations. To do that they presented subjects with pictures from different categories of objects and asked them in separate runs to imagine the objects previously seen. SVM classifiers were trained on fMRI data during perception and imagery conditions, and yielded above chance classification accuracy when tested on the same condition. In order to test the shared neural representations hypothesis they used classifiers trained on perception blocks and tested them on imagery blocks to see if the hyperplanes selected to distinguish between perceived stimuli would correctly classify imagined ones. Very interestingly it was the case, showing that cortical back projections can selectively re-activate patterns of activity underlying perception. This “trick” in the use of classification algorithm to train and test on different experimental conditions is called cross-classification and we have used this technique in order to test shifts of neural representations between categories (see chapter III.). For a review of this method see Kaplan et al., (2015).

MVPA techniques have also been used to build representational spaces in order to link neural and perceptual representations (Kriegeskorte et al., 2008a), individual perceptual differences and their neural representations (Charest et al., 2014) or matching man and monkey categorical object neural representations (Kriegeskorte et al., 2008b). We won't detail these methods because it falls out of the scope of this introduction but see (Haxby et al., 2014) for a review.

The way the VVP represents visual categories has been extensively studied but the way it is carved up during development and with experience is an area of research in full expansion. In the next section we will review what we know about the plasticity of the VVP.

4. Plasticity of the VVP

a. Development

Perceptual functions develop throughout childhood and adolescence (Mondloch et al., 2003). Even visual field maps seem to reach adult like state at 7 years old (Conner et al., 2004). For object selective cortex studies are showing an even slower development with areas associated with faces, objects and places, suggesting that these areas might take longer than EVC to mature. For example ERP studies showed that the latency and amplitude of the N170 was substantially slower and smaller, respectively, for 6 months old infants compared to adults. Golarai et al., (2007) conducted an fMRI study to examine the relationship between perceptual abilities for different visual object categories and the size of their category selective cortex. They observed that the VVP undergoes a maturation process which varies depending on the anatomical part of the VVP studied and that is correlated with the development of category-specific recognition memory.

These results confirm the role of experience in shaping the VVP throughout development (for a review see Grill-Spector et al., (2008))

b. Learning changes neural representations in the VVP

From developmental studies it is pretty straightforward to hypothesize that neural representations studied using MVPA increased sensitivity will be affected by learning. Op de Beeck et al., (2006) investigated the effect of discrimination learning on new object categories in the VVP. They trained subjects to discriminate exemplars of new objects categories during ten days. They recorded the multi-voxel representations of these new objet categories before and after training and showed that these patterns changed due to the discrimination training.

Their findings supported a dynamic view of the VVP where the cortical representations of object categories are distributed and modulated by experience.

Reading ability acquisition

Reading abilities are another kind of visual learning that has been shown to shape the VVP. Written words form a special category of visual stimuli and their processing requires efficiency to read at a correct pace and invariance to recognize letters or numbers regardless of their case, font, size or color. The visual word form area (VWFA) (Cohen et al., 2000) has been shown to be systematically activated when readers are presented with readable words or pseudo-words (letter string similar to words but not forming an existing word) in an alphabet they have learned. It has also been shown that the reading speed was highly correlated to VWFA activation level (Dehaene et al., 2010). Similar effects were observed in the amplitude of ERP response to words with literate and ex-illiterate adults showing an increased lateralized negativity for words compared to illiterate adults at 170ms. This learning effect is even possible during adulthood (Dehaene et al., 2010) showing that there is a sustained plasticity in the VVP to integrate new visual knowledge (see Dehaene et al., (2015) for a review on large-scale and multi-modal neuroimaging changes following reading acquisition and between literate and illiterate adults).

Throughout the VVP many feats of learning have been shown to be possible from childhood to adulthood. These learning induced changes have been observed across the whole VVP and surprisingly even after childhood. It has even been hypothesized that invariance is built by exposure to spatially continuous views of objects through associative mechanisms (Perry et al., 2010). Such associative learning has been observed in an area anatomically related to the VVP: the medial temporal lobe.

C. The Medial Temporal Lobe

At the end of the VVP the MTL, consisting of the parahippocampal (PHC), perirhinal (PRC) and entorhinal (ERC) cortices, the amygdala and the hippocampus, represents the major final output of the visual hierarchy.
In this section we will oversee the major structures of the MTL surrounding the hippocampus, namely the PHC, PRC and ERC, the functions that have been attributed to them and the link they form between perception and memory.

1. Anatomical structures and their connections

The MTL represents a zone of convergence of information, that increases in complexity as it progresses from the PHC and PRC to the ERC and finally to the hippocampus. In fact, PHC and PRC, heavily interconnected, each receive inputs from uni and multimodal sensory neocortical areas (Suzuki and Amaral, 1994).

The **hippocampus** is a remarkable structure in the brain and is observable in all mammals from hedgehogs to humans. Its shape, often compared to a seahorse, or hippocampus in Latin, gave it its name. The hippocampus is located on the internal face of the temporal lobes. Its phylogenetic origin is older than the neocortex and its cortical tissue, the allocortex, is composed of 3 layers only, contrary to the 6 layers of the neocortex. The hippocampus is composed of three zones: the subiculum, the cornu Ammonis fields (CA1-4) and the dentate gyrus (DG). Surrounding the hippocampus, multiple structures form the paths from different regions of the brain to the MTL: **parahippocampal cortex**, **perirhinal cortex and entorhinal cortex.**

Two thirds of PRC inputs come from the ventral visual pathway, from adjacent areas such as TE and TEO in the non-human primate, which correspond to the anterior part of the ventral visual pathway.
The PHC receives input from the dorsal visual pathway, from areas such as the posterior parietal cortex, retrospenial cortex or dorso-lateral prefrontal cortex, but also from the ventral visual pathway.

The ERC receives two thirds of its inputs from these two cortices, each in a specific part: the PRC projects mostly to the lateral part of the ERC and PHC mostly on the medial part of the ERC and in turn the hippocampus receives most of its inputs from the ERC.

2. Discovery of hippocampal functions and MTL memory system

One of the corner stones of the study of mnesic functions and their neural substrates is Scoville and Milner's description of a profound and selective impairment in human memory after bilateral surgical removal of the medial temporal lobe (Scoville and Milner, 1957). The case of Henry Molaison, aka patient H.M, permitted, in a series of studies, to establish the fundamental principle that acquiring new memories is a distinct cerebral function dissociable from other perceptual and cognitive abilities.

Patient H.M suffered from intractable epilepsy and after diagnoses by William Beecher Scoville, a neurosurgeon, about the source of the seizures he had both his MTLs removed surgically as a treatment.

Although the surgery was successful as a treatment to reduce the occurrences of his seizures, H.M developed heavy anterograde amnesia and partial temporally graded retrograde amnesia. Anterograde amnesia is the inability to acquire new memories, whereas retrograde amnesia is the inability to recall memories preceding the lesion, i.e. before the surgery.

One of the discovery that made this clinical case incredibly interesting and influential is the fact that other kinds of memory functions were preserved such as motor learning, e.g. learning to ride a bike, or working memory, e.g. memorizing a list of words for a few minutes. These dissociations between types of memory launched the elaboration of a great amount of models about memory organization.

As we saw before the medial temporal lobe is composed of several structures each with a particular set of inputs and outputs and that can be defined anatomically, but at that time the major divisions of the MTL were between the amygdala, the hippocampus and its surrounding cortex.

From neuropsychological studies it was difficult to draw conclusions on the exact structure that supported memory acquisition and understand what structures in the

MTL were lesioned in H.M. But extensive animal studies on rats and non-human primates allowed to uncover more precisely which parts of the MTL were responsible of different memory impairments.

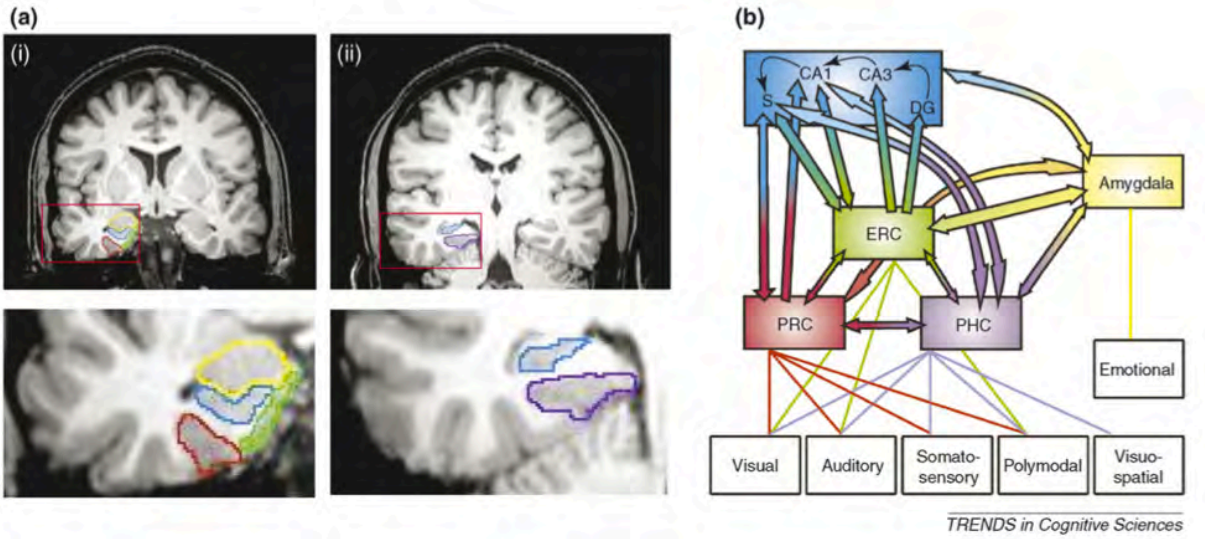

Figure 12. Model of the connections between structures in the MTL. From (Mayes et al., 2007)

The development of animal models in the monkey began with bilateral removal of the MTL that approximated the lesions of patient H.M (Mishkin, 1978). The goal was to damage the whole hippocampus and the amygdala and was termed H^+A^+, H for hippocampus and A for amygdala and the $^+$ sign represented the cortex surrounding these two structures.

In a popular task to probe memory functions subjects are presented with a sample to be remembered and later receive a choice test to evaluate their memory of the sample. The delayed matching-to-sample (DMS) task consists in presenting the picture of an object, the sample, and after a delay present two pictures, one being the sample, subjects are rewarded if they choose the sample amongst the two choices. An alternative version of the DMS task is the delayed nonmatching to sample task (DNMS) in which subjects are rewarded if they choose the other picture.

After H^+A^+ surgeries monkeys were impaired in the DNMS task. Different conditions were tested to investigate which structure was the most critical to

perform this task and after reexamination of histological material of a study with H^+A^+ lesions it was shown that the perirhinal and entorhinal cortices were necessarily damaged when performing the removal with the conventional surgical approach (Squire and Zola-Morgan, 1991). After testing multiple combinations of lesions, from only cortex surrounding the hippocampus (PRC and PHC) to only hippocampus or only amygdala, it was possible to show that not only the amygdala was not a part of the memory system but also that the cortex surrounding the hippocampus was necessary and not only a route to this structure but was also carrying computations which were critical for memory acquisition.

3. Memory and learning types

Another critical finding in the H.M case was that albeit the patient was not able to form new memories, he was still able to access memories formed before the lesion, long-term memories. This and further animal studies performed by Zola-Morgan & Squire (1990) showed that the MTL had a time-limited role in memory. The MTL memory system is indispensable during learning for some types of memory, namely declarative or explicit memory which is divided in two categories: episodic and semantic. Episodic or autobiographical memories are memories of lived events and the context in which they occurred, e.g. last year Christmas dinner, while semantic memory is the memory for facts and concepts, e.g. Toulouse is in southern France.
On the other hand some types of learning and memory are apparently independent from the MTL, for example H.M was still able to acquire new motor skills, i.e. procedural memory.

After a certain period following the acquisition, some memories themselves are not stored in the MTL anymore but probably in other parts of the brain, which would explain why H.M MTL resections left him with unimpaired memories acquired a certain time before the lesion. This process is called **consolidation**: "(The storage of) memories outside the medial temporal lobes by slowly forming direct links between the cortical representations of the experience" - Smith and Kosslyn (2007).

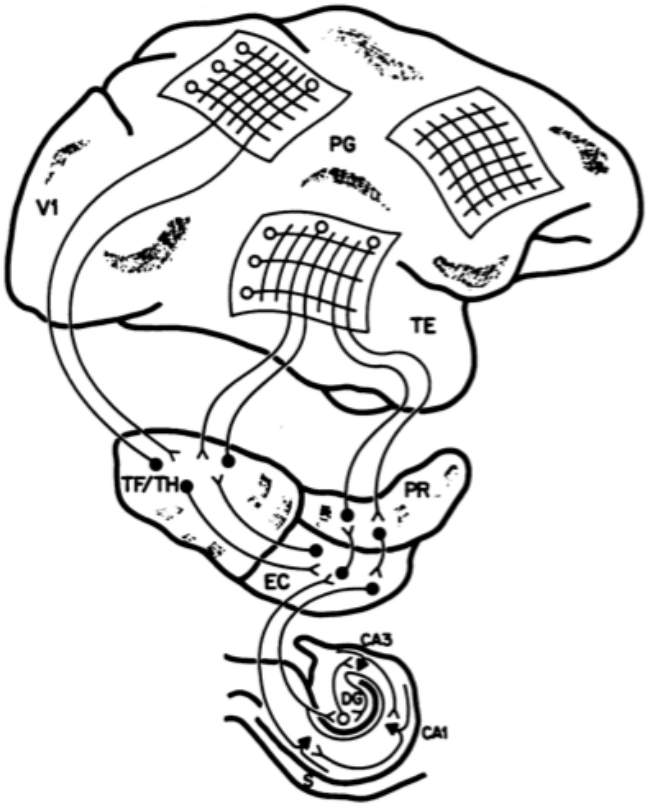

Figure 13. Schematic drawing of primate neocortex plus components of the MTL memory system believed to be important for memory consolidation. This diagram depicts the connections between neocortical sites and MTL structures. The classic consolidation account posits that after memories have been acquired and stored in the hippocampus they undergo a consolidation process. In this process there is a strengthening of the links between remote cortical areas representing the contents forming the memory, by the means of feedback connections from the hippocampus to the ERC and PRC regions in the MTL and finally to neocortical areas. From (Squire and Zola-Morgan, 1991)

4. The MTL: end of the ventral pathway or memory structure on its own?

Role of perirhinal cortex in feature disambiguation

Some of the earliest studies questioning the memory-only function of the MTL were performed by Murray and Gaffan by means of PHC and rhinal lesions, i.e. ERC and PRC. At that time their findings that monkeys could succeed at the DMS task even after rhinal lesions, when stimulus set was very small, and that DMS impairments could be observed in DMS settings with little or no memory demands

challenged the notion that the rhinal cortex was responsible for all aspects of stimulus memory.

Later on, in a series of studies, Buckley and Gaffan howed that PRC lesions led to impaired transfer of discrimination learning between 2D and 3D representations of the same objects and generalization of discrimination learning to new views of familiar objects. Together these findings and others in humans with MTL or hippocampal damage showed that the rhinal cortex has an important role in visual object identification. More specifically the perirhinal cortex is thought to underlie "feature disambiguation" in cases when a subject needs to distinguish between two overlapping stimuli in the feature space, i.e. different views of a face or conjunctions of natural objects

Overall the MTL seems to be involved in perception as well as memory which is coherent from the hierarchical view of the ventral visual pathway that would place the MTL as the latest stages of the hierarchy. Even though these two concepts are often loosely defined in studies trying to disambiguate the two functions, it seems like neither a strictly perceptual or mnesic view of the MTL accounts for the variety of cognitive functions it underlies (for a review see Baxter but also Suzuki or the counter argument on a failure to isolate perceptual and memory demands in the literature).

5. Navigation: a context representation in the hippocampus

Another class of cognitive functions is impaired in the hippocampus lesioned animals and patients: navigation in spatial environments. For example Alzheimer's disease patients show impaired spatial navigation skills and it represents an early sign of the disease (Tangen et al., 2015).

While trying to figure out the nature of hippocampal information processing using neuropsychological and electro-physiological approaches in animals several studies observed a representation of the spatial environment in the hippocampus. More specifically O'Keefe and Dostrovsky (1971) observed that some cells' firing rate increased when a rat was at a particular location in its environment. These

cells were called place cells and demonstrated the role of the hippocampus in spatial adaptation and navigation behavior.

Place cells represent a spatial field through cue association thanks to LTP

One possible molecular mechanism supporting the creation of these spatial receptive fields is the discovery that high-frequency stimulation of hippocampal input fibers can result in long-lasting enhancement of transmission efficacy at downstream synapses, i.e. Long-Term Potentiation (LTP) (Bliss and Lømo, 1973). This mechanism which supports Hebb's synaptic hypothesis for memory storage (Hebb, 2005), could be the way place cells form spatial receptive fields by learning the spatial relationship between cues, e.g. visual cues in the environment. Suzuki et al., (1980) showed exactly that: they found that rotating a set of cues outside a maze where the rat had been navigating did not impair the spatial navigation, on the other hand transposing the cues, which altered the spatial relationship among stimuli, impairs performance significantly.

Thus it seems like the hippocampus acts as an associator of events or cues (Wallenstein et al., 1998).

The spatial receptive fields of place cells have also exhibited a learning process beyond learning to represent the environment through surrounding cues. As Mehta & al. (1997) put it: “Hebbian long-term potentiation predicts that during route learning the spatial firing distributions of hippocampal neurons should enlarge in a direction opposite to the animal's movement”. By recording hippocampal neurons in rats in a closed track paradigm they were able to observe an asymmetric expansion of hippocampal place fields which encoded the cue sequence direction, suggesting a predictive mechanism. This type of anticipation mechanism will be further developed in the next chapter.

Summary and objectives of the thesis

As we have seen, there is many ways in which the different aspects of the visual system can be modulated by experience through temporal and spectral changes, or in the structure of the neural representations of visual objects. Visual learning can take many forms and happen on different timescales, in this thesis we focused on associative learning in pairs of visual stimuli categories or in a visual sequence of visual objects. Specifically we addressed how the brain functional organization integrates this new information at different timescales from hours to weeks of learning.

In the first study we were interested in the impact of learning a visual sequence on the processing stages of the visual system and the mechanisms involved in expectation of an incoming stimuli. We show evidence that the patterns of activity reflect a selective expectation of the next-to-come stimulus and characterized the neural mechanisms using spectral decomposition to show that these patterns were elicited in specific frequency bands of neural activity.

In the long term newly acquired associations are thought to consolidate and re-structure the functional architecture of the relevant neo-cortical sites, i.e. the ones representing the associated perceptual content, in order to be stored in long-term memory (Squire and Zola-Morgan, 1991). To monitor these changes we have conducted fMRI experiment in which subjects underwent two scan sessions separated by 15 learning sessions in which they learned associations between different visual object categories. We compared the multi-voxel patterns of visual category representations before and after this long-term associative learning paradigm. We were able to show a reorganization of the VVP in which the associated categories' multi-voxel representations shift to become more similar. These changes were more pronounced in the posterior fusiform and FFA, two areas that have been closely linked to object perception and object learning. The changes in the representation of object categories were also correlated with the amount of priming associated categories could elicit between one-another.

Finally in the last part of this thesis I summarize the present findings, discuss their limits and relevance to what we currently know about visual learning at different timescales and briefly outline some possible perspectives for future research.

Chapter II: Expectation mechanisms in visual sequence learning

A. Introduction

It has been shown many times that the visual system processes information in a succession of steps that are thought to be, at least to some extent, represented by specific areas in the brain (Van Essen, 2004). When trying to characterize a processing system, and, in the context of this thesis, how it is impacted by experience, it is important to consider the intermediate steps built in it, in other words the different information representations, the timing at which they emerge, the underlying fine temporal mechanisms, e.g. oscillatory components, and which part of this system displays learning effects.

1. Time-course of visual processing

In the context of visual perception the non-human primate brain has been studied for a long time. Here is a canonical description of the timing and anatomo-functional stages of visual processing during the first sweep of the visual hierarchy, from light stimulation in the retina to recognition of a visual stimulus (Thorpe and Fabre-Thorpe, 2001). The first step is retinal stimulation; the signal from retinal ganglion cells is then sent and reaches the Lateral Geniculate Nucleus around 30-50ms later before getting to the primary visual cortex between 40 and 60ms. From there processing continues to intermediate areas of the VVP, e.g. V2 and V4 around 50-70ms and 60-80ms, respectively. It then reaches areas of the posterior infero-temporal cortex (PIT) around 70-90ms, and anterior infero-temporal cortex (AIT) around 80-100ms. These last two areas are the monkey equivalent of category-selective cortex in humans. The infero-temporal cortex then projects to a variety of areas including the MTL, which will be developed in the article and the discussion of this chapter, and the prefrontal cortex (PFC) between 100-130ms. To produce a motor response the signal needs to pass by pre-motor areas and then reach motor areas. This whole process produces a motor response typically around 250ms, i.e. reaction time, for a simple visual categorization task, e.g. animals versus non-animals.

This canonical description of the major temporal steps from retinal stimulation to motor response is summarized in Figure 14.

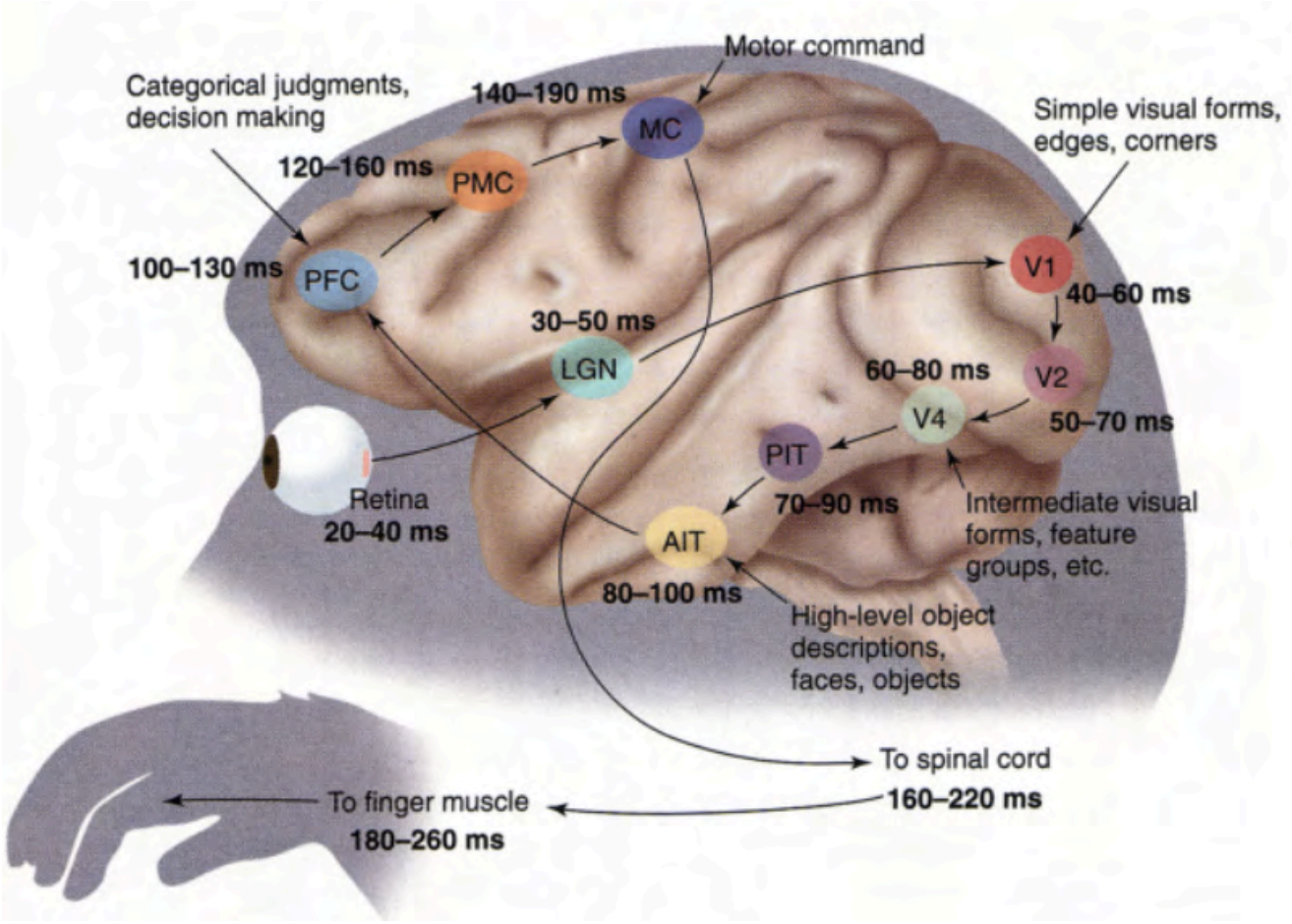

Figure 14. Canonical time-course of visual processing in the monkey's brain for a simple categorization task. From (Thorpe and Fabre-Thorpe, 2001)

But this time-course, which is similar in most ways in humans, can be modified for example through long-term learning to recognize objects such as faces as can be seen in Figure 15.

Electrophysiological studies using non-invasive recording techniques can capture the time course of these processing stages in adults as well as in human infants. It has been used to study differential responses to visual categories and how these develop with experience relative to these categories. For example in adult scalp recordings a negative wave is observable on occipito-parietal electrodes around 170ms, i.e. thus named N170, when contrasting perception of faces with other

object categories. This wave is also sensitive to the inversion and contrast of face stimuli showing that it reflects, at least, some parts of face processing network. In six-month-old infants this evoked potential is also observed for faces more than objects but it is smaller in amplitude and longer in latency (de Haan et al., 2002). This result suggests a process of specialization and maturation in development for visual categories perception. This has also been observed for reading acquisition in a study comparing three adult populations: literate, illiterate and ex-literate (people who acquired reading in adulthood) where it was reported that reading abilities were correlated with enhanced early visual processing in terms of timing, invariance and magnitude of the neural response (Pegado et al., 2014).

Learning thus shapes the time course of visual processing during development from childhood to adulthood as well as during adulthood.

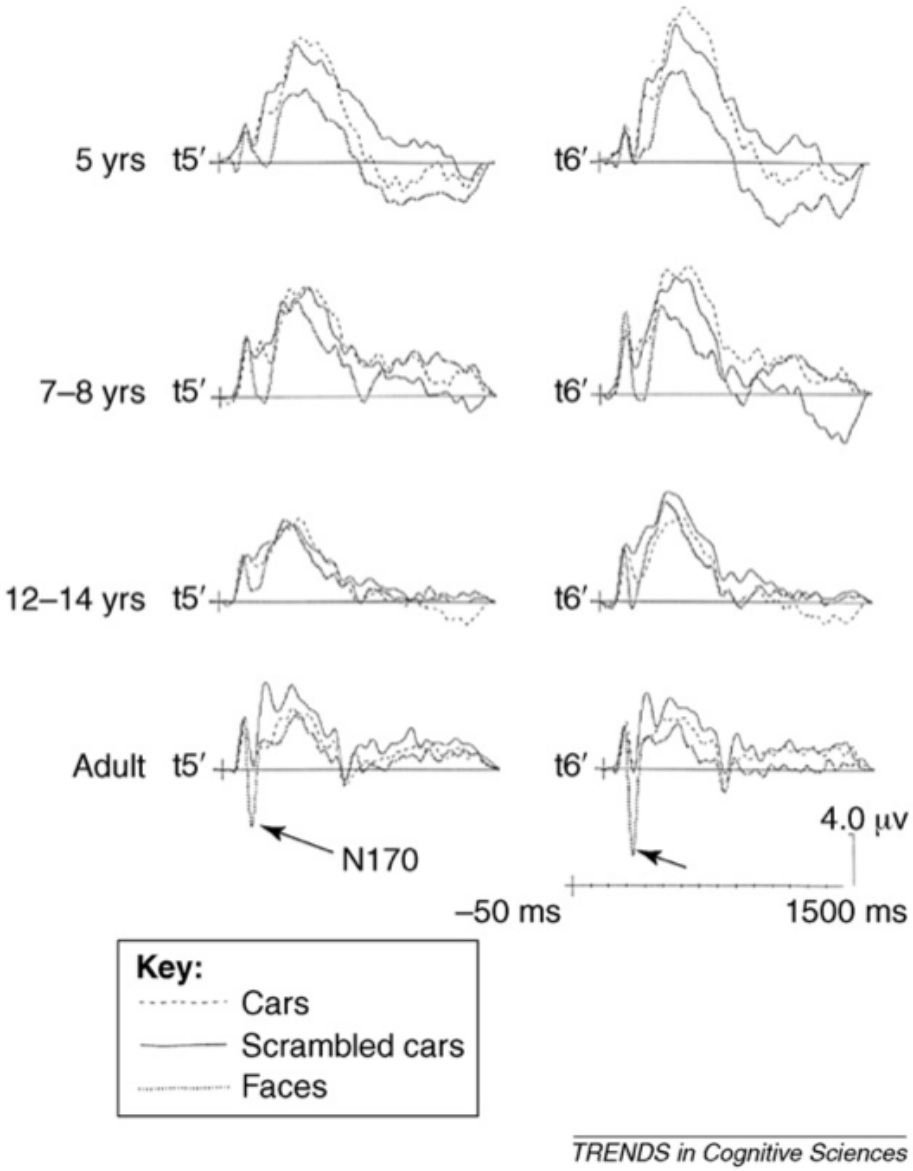

Figure 15. Changes in ERP induced by perception of visual categories. From (Grill-Spector et al., 2008)

Electro-encephalography is a powerful technique that allow us to study these neural signatures of the timing of visual processing as well as the nature of the neural representations underlying perception by contrasting conditions, e.g. inverted or upright faces, or scripts. But what is electro-encephalography?

2. What is electro-encephalography?

a. The technique

The story of electroencephalography begins with biologist Richard Caton who, in 1875, detected in monkeys and rabbits "the presence of electrical currents on the

surface of the brain evidenced by oscillations of the galvanometer needle." Hans Berger applied this technique to humans and the first EEG recorded in 1929 in the form of "permanent potential variations recorded through non-polarizable electrodes applied to a cranial gap" (skull fracture sequel leaving parts of the brain without bone protection) or "on the surface of the intact skull" (Berger, 1929). It is interesting that actually little did Berger care for the electrical phenomenon in itself. After having a premonitory dream, his goal was to materialize "psychic energy" of the human being to identify the "holder of thought" or study its "interpersonal transmission". As a result of this rather esoteric inquiry Hans Berger laid the foundations of electroencephalography in humans.

In recent years, with the advent of computers, the recording paper is replaced by digital recording. Current EEG amplifiers can handle many channels (up to 512 for some devices). Recordings are usually made using Ag/AgCl electrodes (Silver / Silver chloride) of about 1 cm in diameter, most of the time placed according to standard positioning systems, but these positions can be modified in order to increase the sampling density in certain parts of the brain, e.g. more electrodes around occipito-parietal sites in order to record visual potentials more precisely.

b. Origins of the EEG signal

The diverse electrical manifestations captured on the scalp using EEG are most likely the product of pyramidal cells in layer V and III of the cortical sheet (Olejniczak, 2006).

These cells, which are perpendicular to the cortical surface, produce post-synaptic currents oscillating constantly between the soma and their dendrites. Thanks to the ionic charge between different parts of the cell, these neurons act as electromagnetic dipoles. This is illustrated in Figure 16.

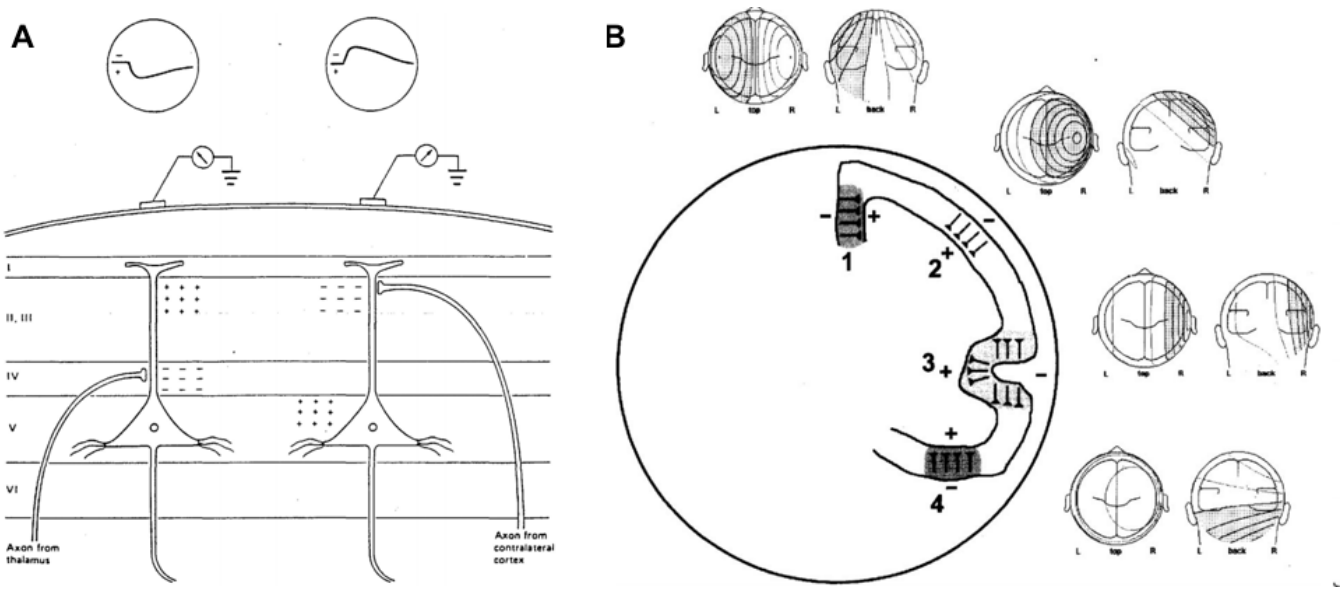

Figure 16. Origins of the EEG signal. (A) Generation of voltage fields as recorded from scalp electrodes from different neurophysiological sources. These two electrodes show opposite effects even though getting the signal from similar neurons because of the location of the synaptic contacts that create opposite dipoles. (B) Schematic of a brain section (coronal) with differently oriented dipoles and their manifestation on the scalp EEG. From (Olejniczak, 2006).

Determining the exact source of the signal captured by the EEG electrode in terms of neural localization is never a straightforward problem because many different dipoles from very distant parts of the brain may contribute to the signal recorded at a single electrode. And if the problem was not complicated enough, the orientation of the particular dipole because of the circumvolutions of the cortical sheet can change the location on the scalp where this dipole will be recorded. It is thus always a controversial topic to infer the location of the neural generators that are the source of the signal recorded at a certain electrode and these source reconstruction methods are still debated inside the field (see Grech et al., (2008) for a review on solving this inverse problem). EEG thus provides maybe the most easy to use neuroimaging technique (less expensive and with relatively simple hardware set up) that can be used to study electrophysiological mechanisms of the brain at a fine temporal scale, but does not provide straightforward information on the topographical origins of the recorded signal.

c. Event Related Potential

One of the conventional methods for EEG analysis is the analysis of evoked potentials. An event related potential (ERP) is a neural signal that reflects the coordinated electrical activity of a recorded set of neurons on the surface of the scalp following the presentation of a stimulus to the subject. The ERP provides a window into the dynamics of network activity linked to various cognitive processes, both at the mesoscopic and macroscopic level, on a time scale that is comparable to the activity of a single neuron, i.e. EEG is typically recorded at a sampling rate of 1000Hz. ERPs come from synchronous interactions within large groups of neurons. This includes very local and very dense interactions at excitatory pyramidal cells and the inter-inhibitory neurons (Olejniczak, 2006).

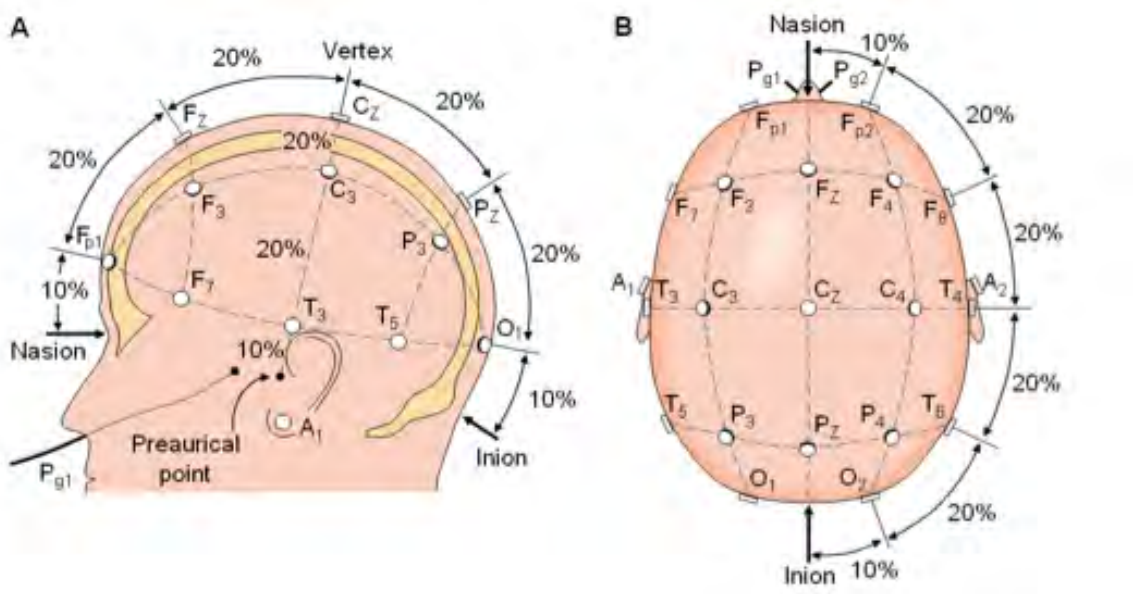

Figure 17. The 10-20 international system of EEG positioning. It is the standard of electrode naming and positioning of the EEG setup on the scalp. This international system is based on iterative subdivision of arcs on the scalp going from posterior to anterior and from side to side of the skull. It uses craniometrical landmarks as reference points: Nasion, Inion; left and right preaurical point. The intersection between these longitudinal and lateral axes is named the Vertex. Electrodes are named according to the underlying lobes: P for Parietal, O for Occipital, T for temporal and F for Frontal. Electrodes near the forehead are named Fp for Fronto-polar. Odd and even electrode numbers represent left and right side of the head respectively. The electrodes are placed at even polar distances around the head as

indicated by the percentages of half circle of each hemisphere. From (Malmivuo and Plonsey, 1995)

d. Spectral analysis of EEG signals

In addition to the ERPs corresponding to transient fluctuations due to coordinated activity from populations of neurons, there is another way of analyzing EEG activity: spectral decomposition. Using signal processing techniques such as Fourier or wavelet decomposition it is possible to estimate another property of the neural signal, which is the speed and amplitude at which it oscillates. Oscillations are defined as repetitive variations of a measure, usually in time, around a central value or between different states. For example the cycles of day and night, or the earth's revolution around the sun can be seen as oscillatory processes. We distinguish different oscillations as a function of frequency, that is to say the speed with which the fluctuation in the measure, e.g. the electric potential captured with EEG, occurs. Figure 18 recalls the main EEG rhythms studied in humans and the corresponding frequency bands.

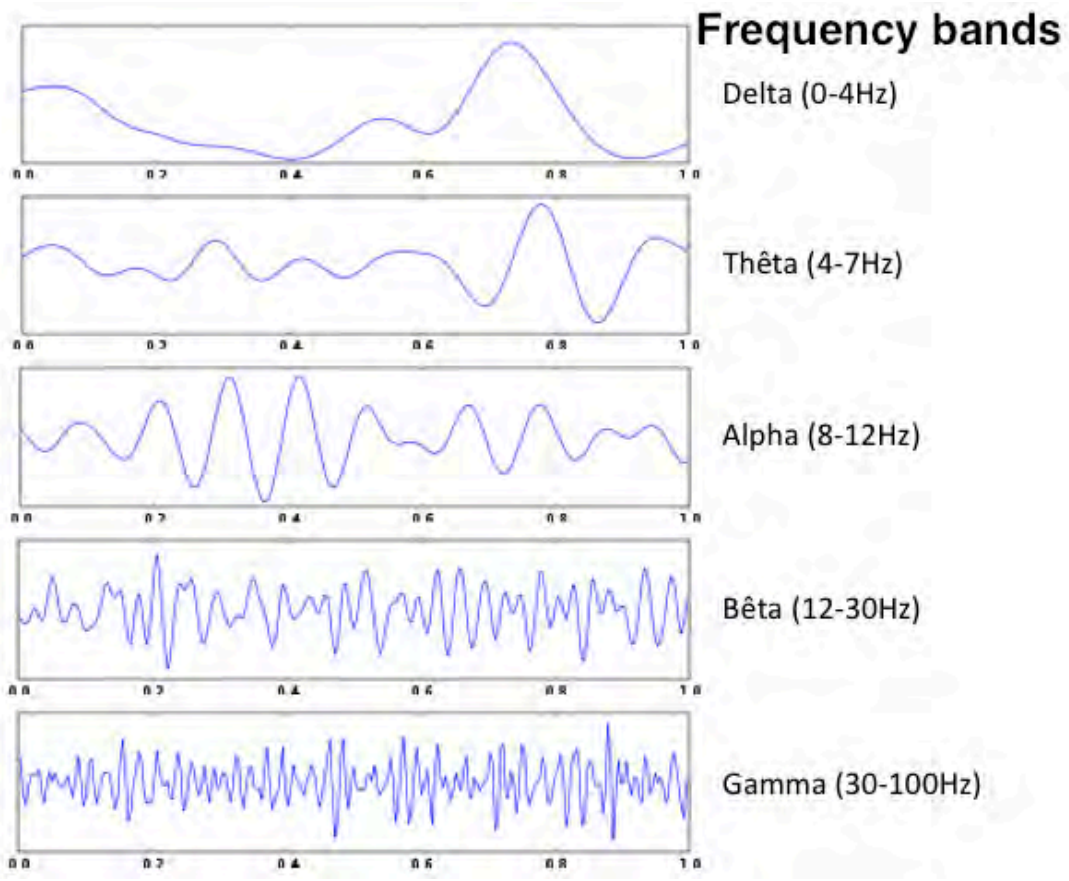

Figure 18. Depiction of the classic nomenclature of frequency bands in EEG. From (Grandchamp, 2012)

3. Remembering sequences of events: a fundamental ability

We learn sequences of event all the time that either form our episodic memory, when they are encountered once and remain in memory as a unique episode, or, when repeated, they form our semantic memory, by factoring out the specific context of exposition and only remembering the relationships between items in an "allocentric" manner. Our experience is perceived through sequences of events and remembering them allows us to predict future outcomes. Because it is such a major format of our experience it is fundamental to understand how it is encoded in the brain and how neural activity shapes our mind to experience this representation of experienced events.

Initial evidence that the hippocampus represents sequence order was found by discovering place cells and the manner in which they replay recently learned or

contextually important information in the same order as experienced during learning.

a. Place cells

In the context of spatial navigation the brain uses external sensory cues, e.g. visual landmark, smell, sound, to create a mental map of the uncharted environment. The neural basis of navigation have been studied for decades and earned some of their most prominent investigators a Nobel Prize in 2014. Certain hippocampal cells described initially by their unique functional property to fire when the animal is at a particular location in the environment are called **place cells** (O'Keefe and Dostrovsky, 1971).

As some neurons in the visual system fire when their optimal stimulus falls in a certain part of the visual field, their receptive field, place cells fire when the animal enters a limited portion of space: their place field. Studies in rats have shown anticipatory modulation of these place fields through experience (Mehta et al., 1997). During the first pass of a rat through a maze, place cells in the hippocampus start to create a representation of that environment by acquiring selectivity to different locations, i.e. place fields. These fields can be modeled at first as a 2-dimensional Gaussian, representing the firing rate of the place cell as a function of spatial location, see Figure 19.

But as the rat passes through certain locations in the maze, thus activating some place cells in a sequential manner, thus learning to navigate in the maze, a place cell will start to fire progressively earlier as the animal progresses towards the location represented by its place field. There is thus a skewing of the place field in the direction opposite to the rat's motion, representing an anticipatory mechanism. In the 2D Gaussian model of place fields this represent a skewing of the Gaussian towards locations visited just before the rat enters this particular place field, this is pictured in Figure 20. Place cells space representation is critically dependent on external sensory cues like the size of the box the rat is placed in, the position of certain landmarks, etc. (Moser et al., 2008).

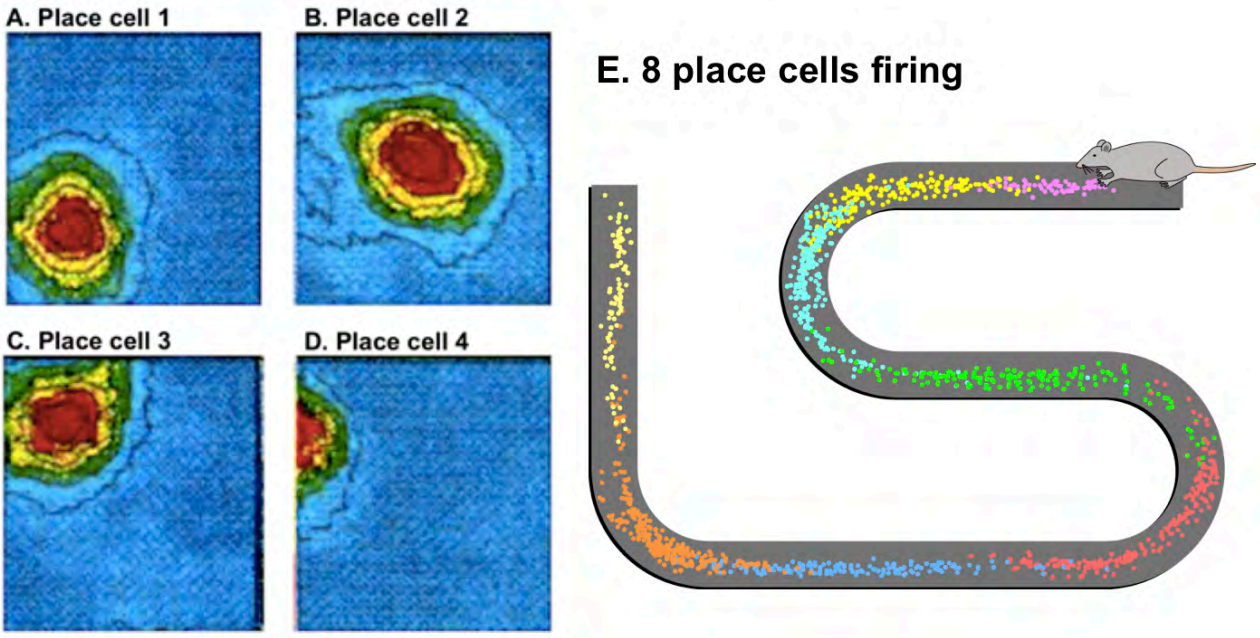

Figure 19. Place fields and place cell firing. (A-D) Four place fields from four different place cells in a rectangular environment. These color map represent firing rate for four different place cells recorded from a rat's hippocampus while placed inside a rectangular box. The colors of these plots represent the firing rate from blue to red for lower to higher neuronal firing rate. From (Wilson and McNaughton, 1993) (E) Eight different place cells firing. Each colored dot represents a hypothetical place cell action potential. As is pictured each place cell fires when the rat is located at the particular position on the track which represents its place field. From Wikipedia (2015)

Place cells are found in CA1, a sub-region of the hippocampus, which receives inputs from multiple sub-regions of the medial temporal lobes (Mayes et al., 2007). These upstream areas are thought to gather sensory information and create invariant unimodal perceptual representations and multi-modal spatial representations, e.g. distances, limits of the environment, for example some rodent studies it was the limits of the box the rat was . Associative learning mechanisms provide an enticing way of generating anticipatory skewing of place fields, mediated through links between sensory cues used as landmarks to navigate through the environment.

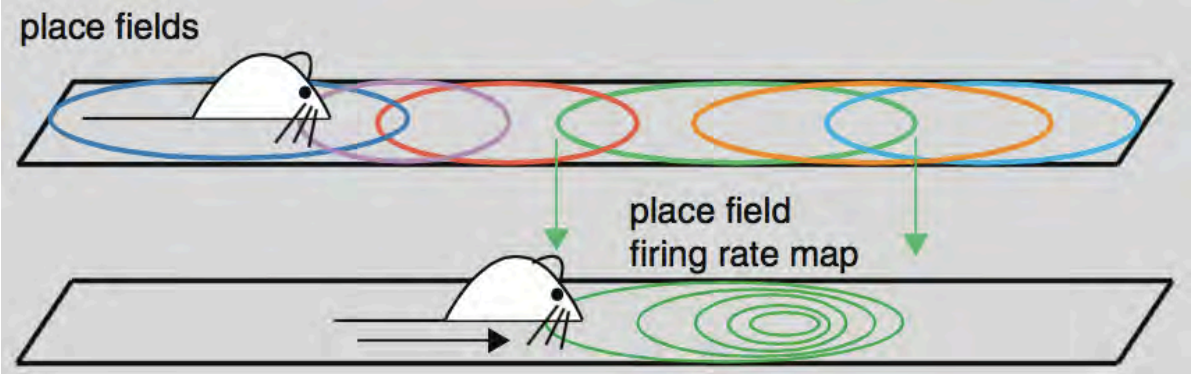

Figure 20. Anticipatory skewing of place fields. On the top schema place fields of a rat navigating in a path, represented by the black rectangle, are represented by colored outlines, each color being a different place field. On the bottom schema one particular place field is represented and the anticipatory skewing is represented by the multiple ovals representing the skewing of the place field in the direction opposite to the rat's motion. From (Foster and Knierim, 2012)

How does the MTL connect temporally disparate items? Here are a few theories:

1. One theoretical proposal of how the hippocampus bridges temporally disparate events into coherent, bounded memories is context sensitive cells that develop from the recurrent network of CA3 neurons (Wallenstein et al., 1998). In a nutshell the idea is that cells coding disparate items in the sequence would be bound to the same cell in CA3 that are supposed to hold a representation of the temporal and spatial context, resulting in a indirect associations between items which appeared in a certain time period (span a temporal delay).

2. Time cells: some cells in the MTL change their firing rate slowly thereby providing a way to lump temporally close items together by serving as a background context (Eichenbaum, 2014). Recently it was shown that the population activity of some cells in CA1 changes gradually over time, these might be the "time cells" (Manns et al., 2007).

3. Another proposal is that items get associated through hippocampal oscillations (Jensen and Lisman, 2005). A major physiological characteristic of the hippocampus is the generation of a theta rhythm. This is observable with electrophysiological techniques where a slow oscillation, around 7Hz,

is observed in the population activity. This account of how the hippocampus connects unrelated items encountered in close temporal proximity similar to the previous theory but instead of considering the existence en time cells that would code periods it considers network dynamics that spontaneously generate slow oscillations. This theory proposes that each theta cycle creates a buffer of temporally coincident items (Lisman and Redish, 2009).

All of these theories rely on quick and automatic associations created between temporally related events. These associative mechanisms have been studied outside of the field of spatial navigation and sequence learning by measuring how the hippocampus associate pairs of events, e.g. visual stimuli, together.

b. Associative learning: from single cells to large-scale patterns of activity

First evidence of such associative mechanisms has been described in monkeys in a seminal study by Miyashita (1988) where it was shown that fractal images presented in a fixed sequence created long-term associative memories in the MTL and nearby VVP. Monkeys were presented with a sequence of fractal stimuli many times in which some of them were always paired in the same order during the training procedure for a delayed matching-to-sample task. In an unexpected result, stimuli that were next to each other in the sequence evoked similar neural responses in perirhinal cortex and area TE, which is part of the monkey VVP, after repeated exposure to this stimulus presentation order. This result was thus one of the first evidence of long-term associative learning in MTL neurons and the nearby areas of the VVP.

A recent study by Reddy et al., (2015) has shown a similar associative learning mechanism, although in a short-term associative learning task, in the human MTL (hippocampus and parahippocampal cortex). While patients were implanted with intracranial electrodes during epilepsy screening protocol they participated in a short visual sequence-learning task. This task consisted of learning a sequence of 5-7 images that were presented in a predictable order. After only 11 loops through this fixed sequence, some MTL neurons, which were sensitive to a particular image, started to fire in anticipation of the presentation of their preferred stimulus. This result bridges associative mechanisms from Miyashita's team and anticipation mechanisms in place cells discussed above. It shows visually responsive cells that

start to respond in an anticipated fashion to their optimal stimulus suggesting that place field's anticipatory skewing might be realized by associating stimuli that shape the two neighboring place fields. This general associative mechanism of the hippocampus has been speculated multiple times in the past (Wallenstein et al., 1998).

This phenomenon has also been studied in healthy adults using non-invasive neuroimaging techniques such as functional MRI. Schapiro et al., (2012) have used multivariate pattern analysis (MVPA) techniques to probe the effect of implicit learning of visual pairs. In their study subjects were presented with continuous stream of fractal images while recording their brain activity using fMRI. They were able to show that MTL structures encode statistical regularities in sequence of image presentations (some images were always presented consecutively, while others were only paired in the presentation order one third of the time), without subjects noticing any structure in the sequence. They showed that, after exposure to the sequence, correlation of multi-voxel patterns elicited by images belonging to a regularly co-occurring pair increased compared to correlation of patterns of non-associated fractals. These results could be a large-scale population-level correlate of the same process observed by Miyashita and colleagues at the neuronal level.

4. Multivariate analysis of time-resolved signals

Using MVPA techniques, such as classification using SVMs, on time-resolved signals to study the temporal dynamics of neural representations is a new exciting field that is currently popular in vision, memory and other cognitive functions. It has been used to uncover object location coding (Chakravarthi et al., 2014), dynamics of invariant object recognition (Isik et al., 2014) or periodic replay in working-memory maintenance (Fuentemilla et al., 2010). This approach allows to monitor the "onset" and maintenance of neural representations, that can be gleaned from patterns of activity recorded through MEG/EEG, and to study the role of different spectral components in these processes. It has been shown that the magnitude of classification performance on scalp topographies between different experimental conditions is correlated with behavioral measures such as reaction times

(Chakravarthi et al., 2014). Therefore these measures are informative on how the timing of neural representations leads to our perceptual experience.

We will therefore use these techniques to study if visual sequence learning induces selective patterns of activity in the absence of an expected event

In the study presented in the next pages we wanted to examine if in the absence of an expected event sequence learning mechanisms selectively evoke representations of this event?

To address this question we realized a sequence learning study similar to Reddy et al., (2015), presented earlier in the introduction of this chapter, but added a new condition in which in 50% of the trials instead of an image appearing only a gray square came onscreen. Such a design creating expectations and the absence of the predicted sensory input has been recently realized using fMRI. Kok et al., (2014) recorded activity in the visual cortex while participants were expecting a visual pattern to appear. Using MVPA techniques they were able to show that prior expectations will trigger the formation of a neural template in sensory areas most probably in order to process incoming sensory stimulation more efficiently. We thus wanted to characterize the fine temporal mechanisms, e.g. spectral components, which were involved in this process of sensory template formation in the context of stimulus expectation.

As will be presented in the next part (B.) subjects were instructed to learn a sequence of images that was continuously presented while in half of the trials the image that should appear in the sequence was omitted and replaced by a gray square. We hypothesized that as a result of learning the sequence, brain activity in the absence of the expected stimulus should reflect the identity of the event that should have been presented at that point in the sequence; alternately, if no learning occurred activity patterns during these trials should not be distinguishable.

B. Paper (in preparation)

Electrophysiological predictors of an expected stimulus during visual sequence learning

Mehdi Senoussi [1, 2], Leila Reddy [1, 2], Rufin VanRullen [1, 2]
1 Université de Toulouse; UPS; Centre de Recherche Cerveau et Cognition ; Toulouse, France 2 CNRS; CerCo; France

Electrophysiological predictors of an expected stimulus during visual sequence learning

Mehdi Senoussi [1, 2], Leila Reddy [1, 2], Rufin VanRullen [1, 2]

1 Université de Toulouse; UPS; Centre de Recherche Cerveau et Cognition; Toulouse, France

2 CNRS; CerCo; France

Abstract:

Learning sequences of events is a fundamental ability, allowing us to remember landmarks on the way to a new job, or to play a song without the partition sheet. Anticipation of upcoming stimuli in a known sequence can improve their detection, but the way our brain builds up this directional association between stimuli is poorly understood. Here we investigated whether learning a visual sequence would elicit neural activity patterns selective for the next-to-come stimulus. Sixteen human subjects performed a visual sequence-learning task while their neural activity was recorded using Electroencephalography (EEG). Six images from distinct categories (car, pinecone, face, camel, house and phone) were presented in a predictable sequence, as if rotating on a virtual "wheel" composed of 6 black square canvases. In each trial the wheel rotated to the next position for 0.5s, stopped and the central black canvas revealed either the corresponding image (Stim trials, 45% of all trials) or a neutral gray square (Catch trials, 45% of all trials) for 1s. Then the image turned black again, and the wheel started rotating for the next trial. In randomly interleaved test trials (10% of all trials) the sequence was stopped and subjects reported via button press which of two simultaneously displayed images was to appear next. All images were equalized in 2D Fourier power spectrum. We hypothesized that as a result of learning, brain activity in Catch trials (with no image on the screen) should reflect the identity of the image that should have been presented at that point in the sequence; alternately, if no learning occurred all catch trials should be treated similarly by the brain. To uncover selective brain activity we used a Support Vector Machine classifier based on spatial patterns of spectral information (for each time point and oscillatory frequency, the amplitudes across 64 electrodes). To limit the influence of residual activity from preceding Stim trials on the classifier performance, we only considered trials preceded by a catch trial. As a first validation of our analysis, we computed classification accuracy during Stim trials (image on-screen). On average across subjects, the classifier reached 62% accuracy, with chance level at 16.6% (permutation test, 100,000 surrogates, $p < 10^{-5}$). Critically, on Catch trials we could also decode above chance the category of the image that should have appeared (20.5 %, $p < 10^{-6}$), even though only a neutral gray square was on the screen. This ability was mainly driven by high-alpha or low-beta frequencies (12-17Hz). These results show that learning a visual sequence induces selective oscillatory activity in the absence of an expected stimulus.

INTRODUCTION

Learning sequences of events is a fundamental ability that we use on a daily basis: from remembering visual landmarks when traveling in a new city to remembering a new song by hearing it once. The ability to learn and remember a relationship between initially unrelated items is termed associative memory. Studies have shown that in the example of pairs of stimuli, associative learning leads to automatic re-activation when presented with a subset of the originally associated items. For example it has been shown that implicit associative learning in the context of a sequential visual presentation containing visual contingencies evoked increased hippocampal activity for predictive stimuli and triggered perceptual anticipation which induces predictive potentiation, i.e. increase in activity due to prediction of stimulus appearance, of category selective ventral visual cortex (Turk-Browne et al., 2010).

In the context of spatial navigation the brain uses external sensory cues, e.g. visual landmark, smell, sound, to create a mental map in order to be able to navigate efficiently this new environment. The neural basis of navigation have been studied for decades and it has been shown that some hippocampal cells had a unique functional property to fire when the animal was at a particular location in the environment, these cells were named **place cells** (O'Keefe and Dostrovsky, 1971). Similarly to visual neurons, place cells fire when the animal enters a limited portion of space: their place field. Studies in rats have shown anticipatory modulation of place fields through experience (Mehta et al., 1997): as the rat passes through a place field over and over again, thus learning the relationship between different parts of the environment, place field show a skewing in the direction opposite to the rat's motion, representing an anticipatory mechanism. Place cells space representation is critically dependent on external sensory cues like the size of the box the rat is placed in, the position of certain landmarks, etc. (Moser et al., 2008).

Place cells are found in CA1, a sub-region of the hippocampus, which receives inputs from multiple sub-regions of the medial temporal lobes (Mayes et al., 2007). These upstream areas are thought to gather sensory information and create invariant unimodal perceptual representations and multi-modal spatial representations, e.g. distances, limits of the environment. Associative learning mechanisms provide an enticing way of generating anticipatory skewing of place fields, mediated through links between sensory cues used as landmarks to navigate through the environment as has been proposed by Wallenstein et al., (1998).

First evidence of such associative mechanisms has been described in monkeys in a seminal study by Miyashita (1988) where it was shown that fractal images presented in a fixed sequence created long-term associative memories in the MTL. Monkeys were presented with a sequence of fractal stimuli many times in which some of them were always paired in the same order during the training procedure for a delayed matching-to-sample task. In an unexpected result, stimuli that were next to each other in the sequence evoked similar neural responses in TE neurons after repeating exposure to this stimulus presentation order. This result was thus one of the first evidence of long-term associative learning in MTL neurons.

A recent study by Reddy & al (2015) has shown a similar associative learning mechanism in humans (while they were implanted with intracranial electrodes during epilepsy screening protocol). Even more interestingly they were able to show that in a very short visual sequence-learning task some MTL neurons started to fire in anticipation of the next-to-come stimulus. This result bridges associative mechanisms from Miyashita's team and anticipation mechanisms in place cells discussed above by showing visually responsive cells that start to respond just before their optimal stimulus is on suggesting that place field's anticipatory skewing might be realized by associating stimuli that shape the two neighboring place fields. This general associative mechanism of the hippocampus has been speculated multiple times in the past (Wallenstein et al., 1998).

Studying MTL functions in associative learning has also been accomplished in healthy humans using fMRI. Schapiro & al (2012) have used multivariate pattern analysis (MVPA) techniques to probe the effect of implicit learning of visual pairs. In their study subjects were presented with continuous stream of fractal images while recording their brain activity using fMRI. They were able to show that MTL structures encode statistical regularities in sequence of image presentations (some images were presented consecutively more often), without subjects noticing any structure in the sequence. They showed that, after exposure to the sequence, correlation of multi-voxel patterns elicited by images belonging to a pair increased compared to correlation of patterns of non-associated fractals. These results could be a large-scale population-level correlate of the same process observed by Miyashita and colleagues at the neuronal level.

Using MVPA techniques on time-resolved signals to study the temporal dynamics of neural representations is a new exciting field that is currently popular in vision, memory and other cognitive functions. It has been used to uncover object location coding (Chakravarthi et al., 2014), dynamics of invariant object recognition (Isik et al., 2014) or periodic replay in working-memory maintenance (Fuentemilla et al., 2010). This approach allows monitoring the appearance of neural representations , at least as recorded using non-invasive electrophysiological techniques, and maintenance at a very fine temporal scale in patterns of activity recorded through MEG/EEG, and to study the role of different spectral components in these processes.

For instance a recent study has investigated the dynamics of neural representations of visual objects' location using EEG and ERPs. The subjects were performing an object discrimination task, discriminating circle versus square stimuli, and the classifier was trained to distinguish between location of the object onscreen based on the distribution of ERP on the scalp, i.e. object presented on the left or right of the fixation point. This result suggests that the dynamics of visual object neural representations as measured with EEG scalp topographies are informative at a behavioral level, i.e. they correlate with reaction times and accuracy.

Another part of the literature has been investigating the effects of expectations in the context of the predictive coding theory (Friston, 2005). A recent study has investigated how prior expectations facilitate perception through top-down processes. In this study Kok and al., (2012) used fMRI and MVPA

techniques to ask whether expectation mechanisms sharpen the representations of expected stimuli found that there was a dissociation between expectation effects on the amount of activity and the amount of information represented an area. Subjects had to perform orientation discrimination and contrast discrimination tasks on Gabor patches and they a decrease in V1 voxels which were sensitive to the orientation of the Gabor that was expected but an increased orientation classification using MVPA techniques. Their result thus shows a sharpening of the neural response for the expected sensory input. The same group then carried another study in which they were able to show that in a similar task where subjects were expecting a certain sensory input, that these prior expectations evoked a feature specific pattern of activation similar to the pattern observed during the actual presentation. These results thus shows that expectations mechanisms can reactivate the neural representations of expected sensory inputs, but it is still unknown what neural mechanisms underlie these processes?

To address this question we performed a sequence learning study similar to the one in Reddy & al (2015), where subjects were presented with a sequence of images and had to learn the order of presentation, but added a new condition in which in half of the trials the stimuli, e.g. face, car, etc., which was suppose to appear at that position in the sequence did not, and a gray square appeared instead. We hypothesized that as a result of learning, brain activity in the absence of the expected stimulus should be selective to the event that was expected; alternately, if no learning occurred activity patterns during these trials should not be distinguishable.

MATERIALS AND METHODS

Subjects and stimuli

Sixteen subjects (age 21-45 years, 5 females) were recruited in this experiment. All had normal or correct-to-normal vision and no history of neurological problems. All subjects provided written informed consent and received monetary compensation for their participation. The local ethics committee approved all procedures

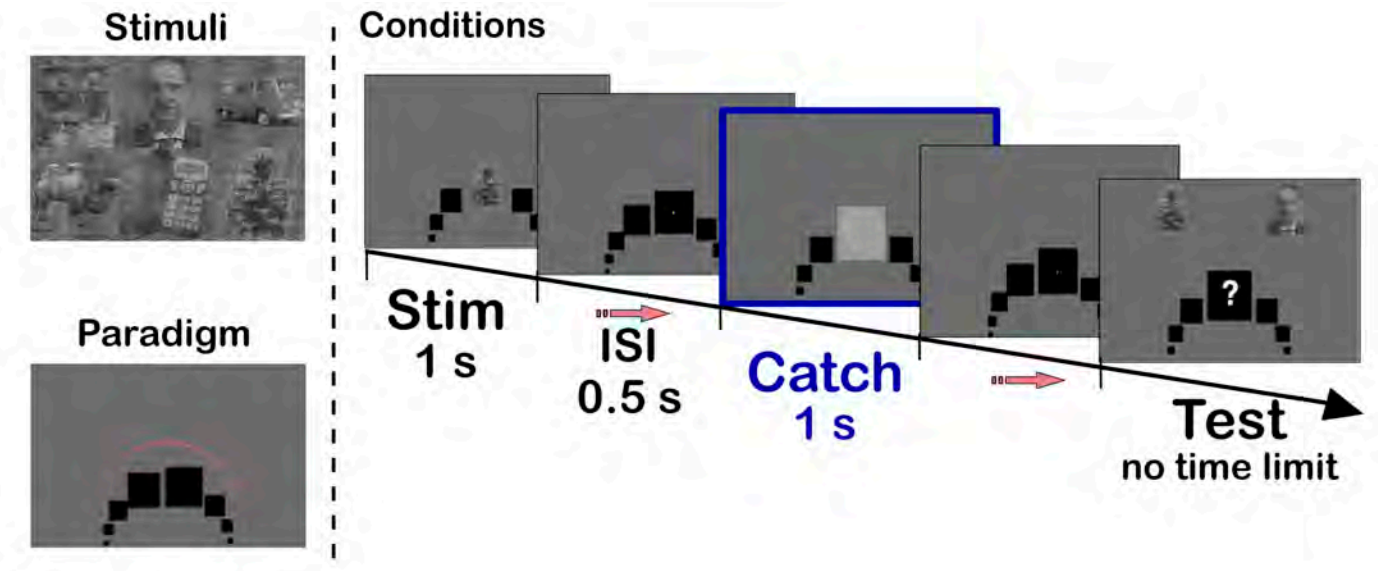

Figure 1. Experimental protocol. Stimuli Six stimuli were gathered from Internet from six different visual object categories. Categories were based on studies on representational similarity analyses which analyzed the distance in neural representations between exemplars belonging to visual object categories such as human faces, animal faces, human bodies, man-made objects, etc. (Carlson et al., 2013; Kriegeskorte et al., 2008). **Paradigm** Stimuli were presented in a predictable sequence as if rotating on a virtual "wheel" composed of black canvases in order to give the sequence a "spatial" aspect as if encountering stimuli or visual landmarks when navigating an environment. Each time a canvas reached the center of the screen it stopped for 1s and revealed one of three trial types. **Conditions** There was three possible trial types: 45% were Stim trials in which the corresponding stimulus at that position in the sequence replaced the black canvas, 45% were Catch trials in which only a gray square replaced the black canvas instead of the corresponding image at that position in the sequence and 10% were Test trials in which a question marked appeared in the black canvas and two of the possible stimuli appeared at the top of the screen prompting the subject to indicate with a button press which of the two should have appeared at that position in the sequence.

Six stimuli from six distinct object categories were gathered from the Internet. The choice of the six visual object categories was based on studies on representational similarity analyses which analyzed the distance in neural representations between exemplars belonging to visual object categories such as human faces, animal faces, human bodies, man-made objects, etc. (Carlson et al., 2013; Kriegeskorte

et al., 2008). This selection was intended to maximize the distance in the multivariate EEG responses to each stimulus in order to maximize our ability to distinguish neural responses, which would be selective to the expected stimuli. The chosen stimuli categories were face, camel, car, house, pinecone and phone. All stimuli were equalized in 2D Fourier power spectrum in order to diminish low-level confounds in the stimuli set. The stimuli were in a square shape and had a size of 7 degrees of visual angle.

Experimental procedure

Stimuli were presented in a predictable sequence as if rotating on a virtual “wheel” composed of black square canvases, see Figure 1. In each trial the wheel rotated to the next position for 500ms, stopped and the central black canvas revealed either the corresponding image (**Stim** trials, 45% of all trials) or a neutral gray square (**Catch** trials, 45% of all trials) for 1s. Then the image turned black again, and the wheel started rotating for the next trial. In randomly interleaved **Test** trials (10% of all trials) the sequence an interrogation mark appeared on the central canvas and two images at the top of the screen and subjects reported via button press which of two simultaneously displayed images was to appear next, the correct answer position (left or right) was randomized. The task of the subjects was thus to learn the order of appearance of each image in order to be able to indicate which image should appear at any moment during the experiment.

The gray image appearing during Catch trials was 50 units in RGB space above the gray background. A fixation cross was present at the center of the central canvas at all time in order to avoid eye movements. The subjects was familiarized with the task and experiment layout by running approximately 10 Catch trials and explaining that either these or the actual image were as likely to appear when the experiment start. The experiment consisted of eight blocks of approximately 10 minutes each. Each block contained 60 loops of the six-images sequence, minus the 10% of Test trials yielded 300 trials per block, hence for a total of 2400 trials for each subject, half of them being Stim trials and the other Catch trials. Subjects were instructed that they could rest their eyes during Test trial and answer as soon as they were ready to resume the sequence. Responses were recorded to measure how well they knew the sequence.

EEG acquisition

64-channel EEG was recorded using a BioSemi Active Two system at a sampling rate of 1024 Hz. A three-channel EOG was also recorded to monitor left horizontal and right horizontal and vertical eye movements and blinks. Data were then downsampled to 256 Hz and epoched from 500ms before trial onset (beginning of ISI rotation) to 1000ms post trial onset (1.5s). Raw EEG time-courses were screened manually on a trial-by-trial basis to reject visible artifacts, eye movements or blinks. Baseline correction was applied by subtracting the average activity between -500 and -400ms relative to stimulus onset for each electrode and trial independently.

EEG pattern classification

To assess whether the scalp distribution in Catch trials was selective of the expected event to come we used classification techniques that try to find distinguishing features between classes of scalp topographies at each time-point, and predict the expected event.

For all classification analyses we used a linear SVM classifier (C=1) and a 10-fold Stratified K-fold cross-validation procedure, both implemented in the Scikit-learn python toolbox (Pedregosa et al., 2011). The input to the classifier was always 10 training examples, which we will refer to as samples, by event (expected stimulus, e.g. face, car, etc.), so 60 samples in total. Each sample was made by averaging approximately 20 single-trials in voltage time-course for the ERP analysis or the amplitude of their complex time-frequency decomposition for the time-frequency analysis. Because each sample was computed arbitrarily (e.g. averaging trials 1 to 10, 11 to 20, etc.) we re-sampled 10 times and re-ran the classification analysis to avoid any sampling bias.

Time-frequency transform was done using Morlet wavelet decomposition implemented in the MNE-Python suite (Gramfort et al., 2013, 2014). The frequencies ranged from 2 to 50hz (linearly spaced by 1 Hz steps) with the number of cycles linearly increasing from 1 to 15 cycles.

Statistical analysis was done using Student T-test and permutation tests based on 100.000 surrogates generated from the re-samples by shuffling labels prior to feeding the data to the classifier.

The Vote Method: Pooling data across classifiers

In order to investigate the amount of information elicited by the presentation of Catch trials (1s) across all time-points/frequencies we designed a new method of classifier output analysis:

- For a given sample a classifier will predict a label (or 'vote') for each time-frequency point
- We then evaluate if the most voted label across time-points (resp. frequencies) is correct
- Finally averaging across all samples yields a new classification score from aggregating across votes

Using the Vote method permits to quantify how selective was the EEG activity across multiple time or frequency points. For example it would allow testing, when aggregating across time-points, if during the 1s Catch presentation neural activity represented the expected stimuli at different moments, as if “building up” the learned expectations for useful behavioral output or to prepare the visual recognition system as it has been shown to enhance the quality of sensory information (Rohenkohl et al., 2012).

RESULTS

Behavioral Performance

Figure 2 shows the average accuracy across all subjects by block. The performances on the 2-AFC task were all above chance (one-sample t-test: $t(15)>8$, $p<10^{-6}$) and there was an increase in performance along blocks (one-way, random-effects ANOVA: $F(7, 105)= 2.73$, $p < 0.02$). A post-hoc Tukey's HSD showed that only block 1 was significantly different from the others ($p < 0.05$), but the classification analysis on the ERPs did not show any difference if trials from block 1 were included or not (one-sample t-test: $p>0.05$). All further analysis thus included EEG data from block 1.

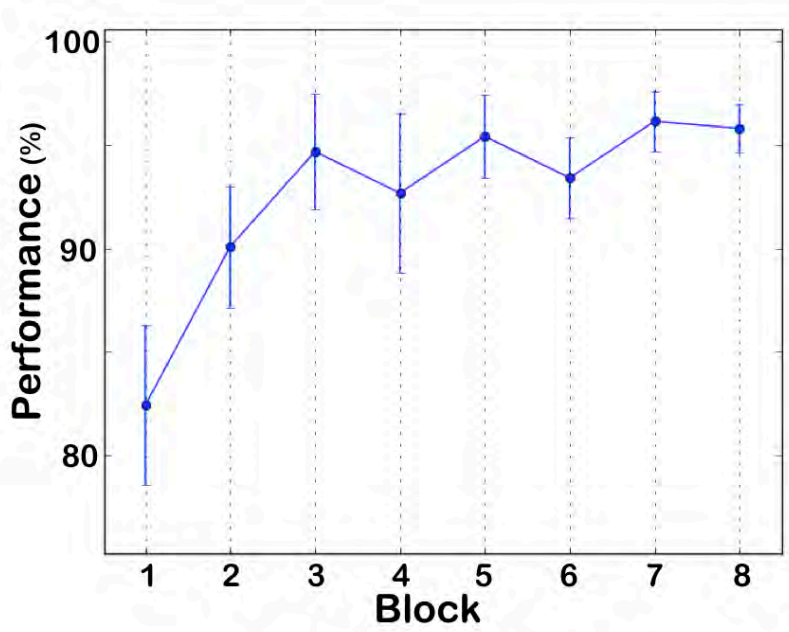

Figure 2. Behavioral performance. Subjects were tested during the session at random moments and had to perform a 2-AFC task by indicating which of two simultaneously presented stimuli should have appeared at that trial. Subjects achieved more than 80% accuracy in the first block (chance at 50%). Their performance kept increasing to reach more than 95% accuracy on the two last blocks. There was an increase in performance along blocks (one-way, random-effects ANOVA: $F(7, 105)= 2.73$, $p < 0.02$). A post-hoc Tukey's HSD showed that only block 1 was significantly different from the others ($p < 0.05$)

ERP preceding trial type control

Because of temporal proximity between trials (500ms ISI) we carried out a control analysis to verify whether the preceding trial type had an effect on the classification accuracies. We thus carried our classification analysis on scalp distribution of samples at each time-point by splitting trials by their preceding trial type, i.e. Stim or Catch, therefore for each trial N, N-1 Stim or N-1 Catch. As is pictured in Figure 3A the classification accuracy is higher for N-1 Stim trials than N-1 Catch trials for trial N Stim or trial N Catch. Classification accuracy is higher for trials N preceded by a Stim (N-1 Stim) in 68% of time-points (47/150) for Stim trials and 52% of time-points in Catch trials (71/150) (two-tailed, paired t-test,

p<0.05). Therefore EEG activity is still carrying information about trial N-1 Stim during trial N presentation. Trial N-1's type influences trial N classification accuracy, we will thus only use trial N-1 Catch.

To ensure that trial N-2 type did not have a similar influence we split trials again by trial N-2 type, i.e. N-2 Stim and N-2 Catch, while keeping trial N-1 Catch. As can be seen in Figure 3B, there is no difference in the classification accuracy time-course between trials whether N-2 was a Stim or a Catch: in trial N Stim there was no time-point where classification accuracy for N-2 Stim and N-2 Catch significantly differed and for trial N Catch only 3.3% of time-points exhibited a higher classification accuracy for N-2 Stim (two-tailed, paired t-test, p<0.05).

All further analysis we thus be carried on trials that were preceded by a Catch (N-1 Catch trials), regardless of the N-2 trial type.

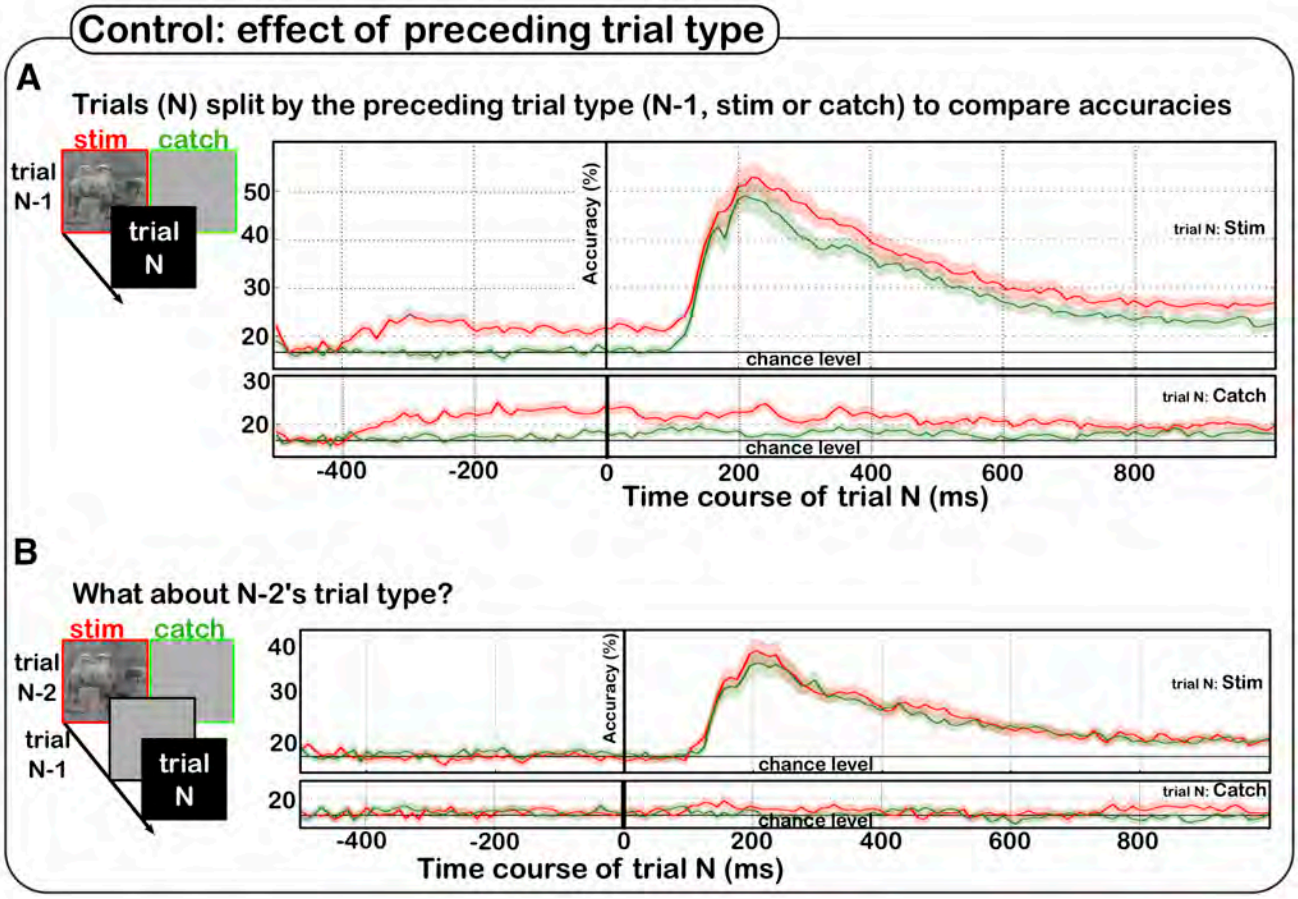

Figure 3. Preceding trial type effect/control. (A) Effect of the preceding trial type (effect of N-1 type on trial N). To control the effect of the preceding trial type we split trials into two groups: the ones preceded by a Stim and the ones preceded by a Catch, and carried out the classification analysis. As can be seen on both plots there are differences in classification accuracy between trials whether they were preceded by a Stim or a Catch trial. This difference is present in both Stim and Catch trials N. Classification accuracy is higher for trials N preceded by a Stim (N-1 Stim) in 68% of time-points (47/150) for Stim trials and 52% of time-points in Catch trials (71/150) (two-tailed, paired t-test, p<0.05). **(B)** Effect of the type of trial preceding the preceding Catch trial (effect of N-2 type on N). To ensure that trial N-2 did not have any effect on trials N preceded by a Catch trial (N-1) we split the

trials N by their N-2 trial type. here is no difference in the classification accuracy time-course between trials whether N-2 was a Stim or a Catch: in trial N Stim there was no time-point where classification accuracy for N-2 Stim and N-2 Catch significantly differed and for trial N Catch only 3.3% of time-points exhibited a higher classification accuracy for N-2 Stim (two-tailed, paired t-test, $p<0.05$).

EEG topographical pattern classification

As concluded in the last section we only used trials preceded by a Catch in this analysis. The classification accuracy time-course showed sustained above chance performance at multiple moments. Early in the pre-trial onset period, from -500 to -400ms the classification accuracy was at chance (16.66%), probably due to the baseline correction done in that time-window. From 400ms to trial onset the classification accuracy rises even though it never reaches significance before trial onset. Post trial onset the classification accuracy continues to rise and reaches 19.95% of classification accuracy which was significant using both t-test (one-sided, paired t-test, $p<5.10^{-4}$) and permutation test ($p<10^{-6}$), see Figure 4. A second consistent peak at 19.7% of accuracy appears at 350ms and is also significant using both t-test (one-sided, paired t-test, $p<0.01$) and permutation test ($p<10^{-6}$). The first peak is comprised in a period of significantly above chance classification from 120 to 180ms (one-sided, paired t-test, $p<0.01$) and permutation test ($p<10^{-6}$).

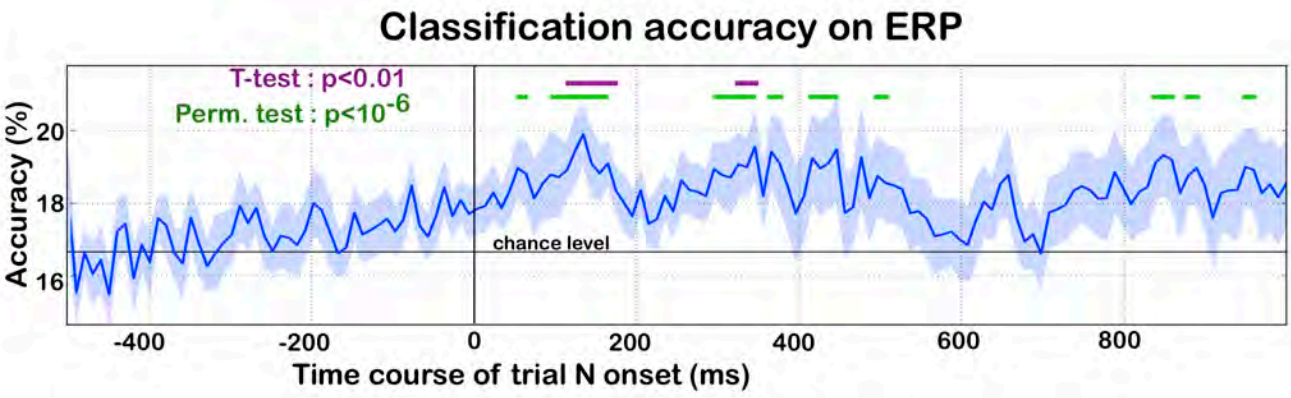

Figure 4. Classification results on N-1 Catch trials.

6-way classification analysis was performed on the EEG samples at each time-point using the 64-electrode ERP scalp map. The classification accuracy is represented in blue and the standard error to the mean across subject in blue shading around the curve. Green and purple lines above the classification accuracy line are results of statistical tests: one-sided Student T-test in purple and permutation test in green. The classifiers reached significance above chance classification accuracy most consistently at two moments: from 120 to 180ms (one-sided, paired t-test, $p<0.01$ and permutation test, $p<10^{-6}$) and from 330 to 360ms (one-sided, paired t-test, $p<0.01$) and in more moments using the permutation test, especially at the end between 850 and 1000ms.

EEG temporal pattern classification

In order to locate the region in which these selective pattern appeared we performed the same classification analysis but instead of using the 64-electrodes distribution of ERPs we used the entire

time-course at each electrode (from -500ms to 1000ms) for the input vector to our classifier. The results show significant classification accuracies in fronto-central, right temporal and bilateral occipito-parietal effects. The fronto-central electrodes show the largest effects with Fz time-course achieving 20.5% of classification accuracy ($p<0.001$).

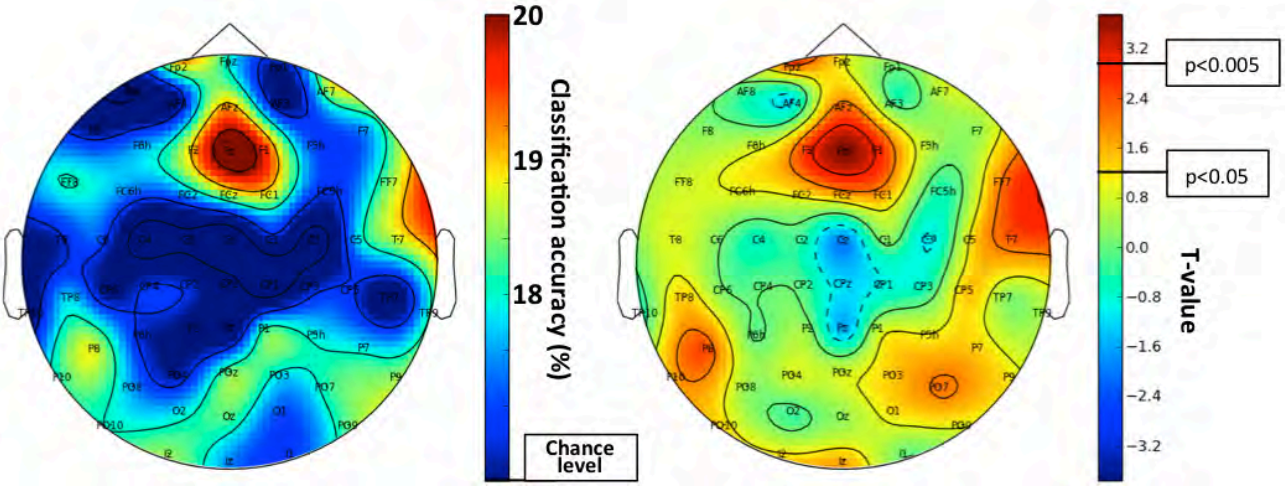

Figure 5. EEG classification on time-course. 6-way classification analysis realized on the time-course by electrode, instead of topographical patterns by time-points. Left: Classification accuracies by electrode. Right: T-values from two-tailed, one-sample t-tests across subjects. Fronto-central electrodes show the largest classification accuracies (20.5%), with Fz being highly significant ($p<0.001$). Some right fronto-temporal electrodes also show significant effects having a classification above 18% and significantly above chance ($p<0.01$). Finally a smaller but significant effect at $p<0.05$ is observed at occipito-parietal electrodes on the right and left sides with classification accuracies reaching 18.5%.

EEG time-frequency classification: time-frequency map

To investigate the frequency specificity of the ERP effects we applied the same method, namely 6-way classification of the expected event at each time-frequency point. This analysis yielded above chance classification accuracies for time-frequency points situated mostly in low-frequency bands.

To investigate the significance of these points we used a cluster test based on surrogates. We generated 100,000 surrogates for each subject time-frequency classification accuracy map by shuffling the labels of the samples fed to the classifier. The shuffling was done once for each batch of resamples to be in the same condition as the original data and avoid inducing more noise. Shuffling the labels at each time-frequency point would have made the structure of the data more variable in a surrogate map than in the real data, thus not comparable.

To get a statistical threshold of cluster size we computed the size of the biggest cluster in the 100,000 surrogates at a certain classification accuracy threshold, thus any remaining cluster bigger than that size in the real data time-frequency classification accuracy map would be significant at $p<10^{-5}$

The remaining significant clusters at a threshold of 18% of classification accuracy are outlined in black in Figure 6. These clusters are mostly in low frequency bands and seem to increase in frequencies from

theta just after trial onset to alpha and then beta frequency ranges towards the end of the trial. A similar result was obtained using a False Discovery Rate (FDR) correction for multiple comparisons threshold.

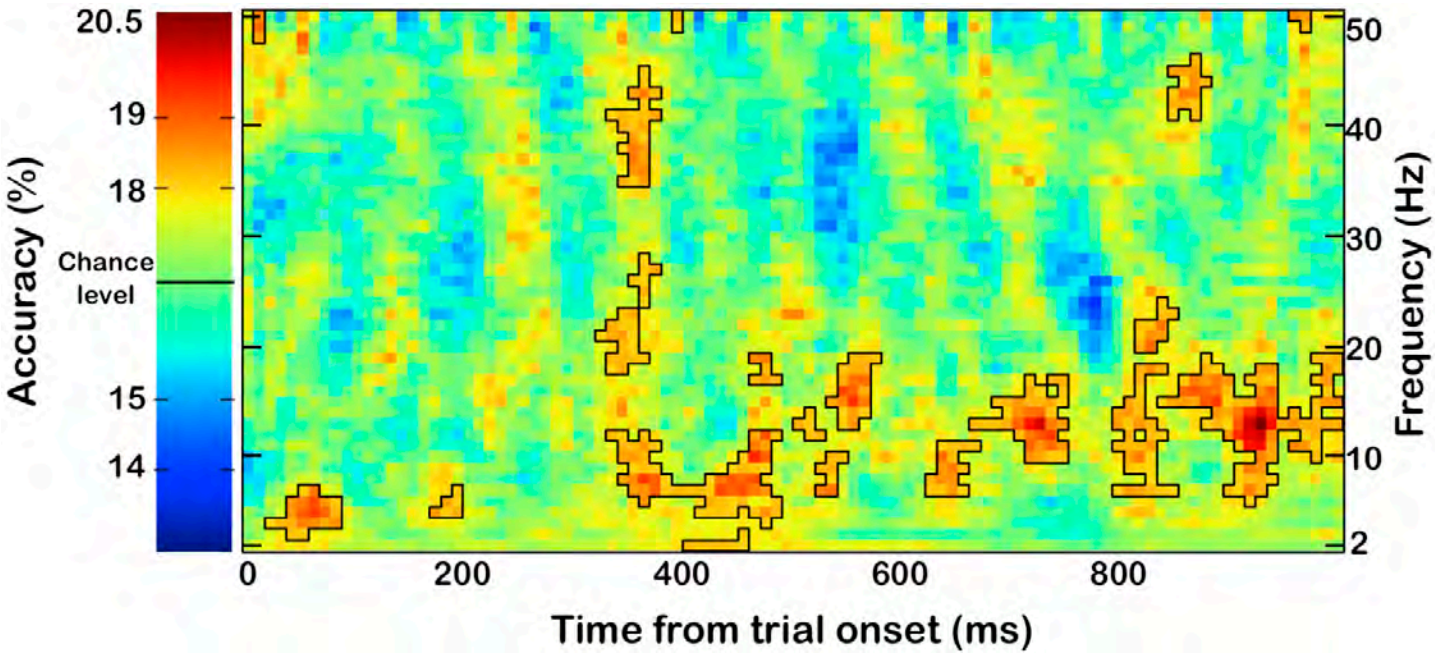

Figure 6. Time-frequency results with cluster test. Classification accuracy at each time-frequency points. Color bar represents classification accuracy from 12.8% (dark blue) to 20.5% (dark red). Black outline represents clusters that were significantly bigger than observed by chance (permutation test, $p<10^{-5}$). When using a False Discovery Rate (FDR) correction for multiple comparisons threshold we get a very similar map with most voxels being significant at $p<0.01$ being comprised inside these clusters.

EEG time-frequency classification: temporal map

In order to measure temporal dynamics of the time-frequency decomposed signal we used two approaches to collapse information across: averaging and the Vote method (see Methods). Simple averaging classification accuracy across frequencies at a certain time-point reduces the classification accuracy because it will take into account frequencies that do not carry any information on the expected stimulus. Using the Vote method allows us to aggregate the information present at any frequency and capitalize that to increase the overall classification accuracy at that time-point, yielding the actual information content across frequencies without uninformative frequencies hampering the overall classification accuracy at a certain time point. Significance was assessed with a one-sided Student T-test for random-effect analysis and a permutation test using 100,000 surrogates for fixed-effect analysis. As can be seen in Figure 7, when averaging across frequencies, classification accuracy reaches significance at different moment during the trial: from 330 to 380ms, 710 to 730ms and 920 to 950ms were significant using a T-test ($p<0.05$). The permutation test ($p<10^{-6}$) revealed more significant points but usually around the same time periods as using the random-effect T-test. Using the Vote method across frequencies, classification accuracy reaches much higher performance, getting to 22% at 350ms.

Above chance classification accuracies were also obtained at different moment during the trial: from 50ms to 70ms, 330 to 380ms and 950 to 970ms were significant using a T-test ($p<0.05$) and using the permutation test ($p<10^{-5}$) a majority of time-points were significant including all time-points significant using the random-effect T-test.

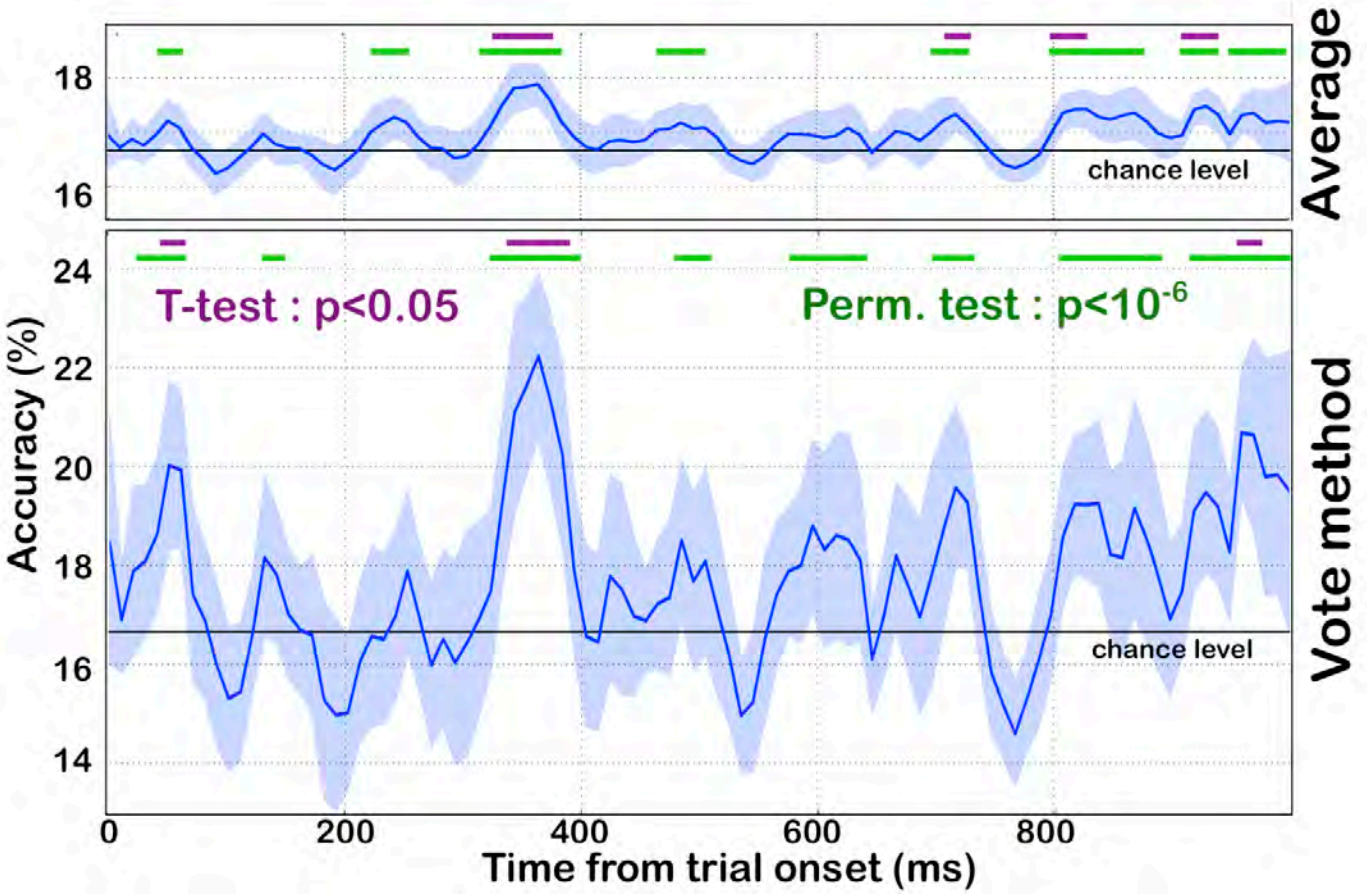

Figure 7. Time-frequency classification collapsed across frequencies.

Axes represent the collapsed classification accuracy of the time-frequency matrix across frequencies using averaging or the Vote method by time-point. **Top**: Average across frequencies. **Bottom**: Vote method across frequencies. Significance was assessed with a one-sided Student T-test for random-effect analysis and a permutation test using 100,000 surrogates for fixed-effect analysis. When averaging across frequencies, classification accuracy reaches significance at different moment during the trial: from 330 to 380ms, 710 to 730ms and 920 to 950ms were significant using a T-test ($p<0.05$). The permutation test ($p<10^{-6}$) revealed more significant points but usually around the same time periods as using the random-effect T-test. Using the Vote method across frequencies, classification accuracy reaches significance at different moment during the trial: from 50ms to 70ms, 330 to 380ms and 950 to 970ms were significant using a T-test ($p<0.05$) and using the permutation test ($p<10^{-5}$) a majority of time-points were significant including all time-points significant using the random-effect T-test.

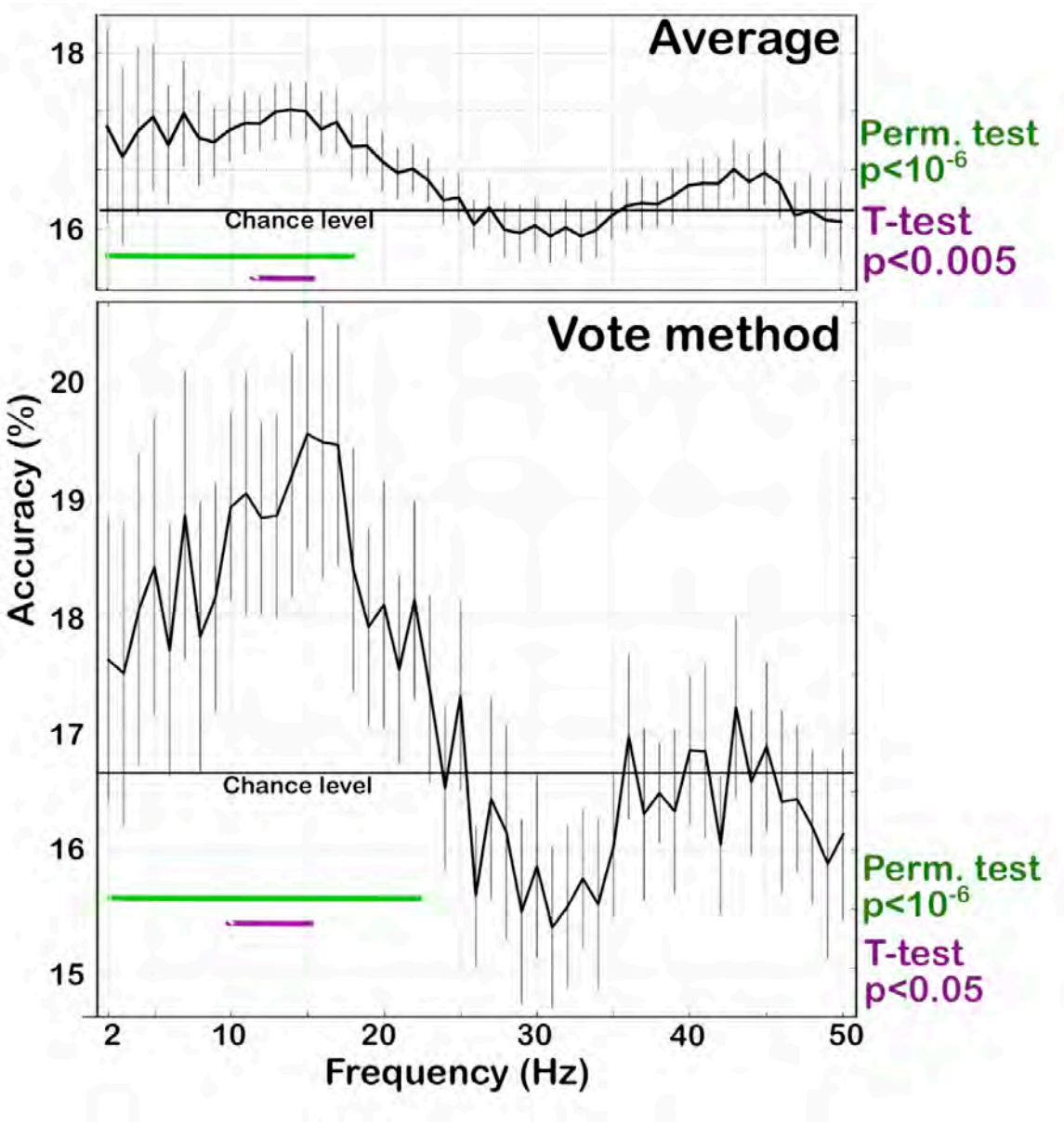

Figure 8. Time-frequency classification results collapsed across time. Axes represent the collapsed classification accuracy of the time-frequency matrix across time-points using averaging or the Vote method by frequency. **Top**: average across time-points. **Bottom**: Vote method across time-points. Significance was evaluated using a one-sided, Student T-test for random-effects analysis and permutation test based on 100,000 surrogates for fixed-effect analysis. When averaging across time, frequencies from 12Hz to 15Hz were significant using a T-test ($p<0.005$) and frequencies from 2Hz to 18Hz were significant using the permutation test ($p<10^{-6}$). Using the Vote method across time, frequencies from 10Hz to 17Hz were significantly above chance using a T-test ($p<0.05$) and from 2Hz to 23Hz using the permutation test ($p<10^{-5}$).

EEG time-frequency classification: spectral profile

As in the temporal profile analysis we investigated the spectral specificity of the expectation effects (Figure 8). Using both the average and vote method across all post-trial onset time points we were able to show that the selective patterns of activity were frequency specific and highly significant across the average and the Vote method and most significant between 12 and 17Hz (one-sided, paired t-test, $p<0.005$ and permutation test, $p<10^{-6}$).

DISCUSSION

Visual learning can take many forms; one of them is sequence learning, which can be considered a subtype of associative learning in that visual stimuli are directionally associated by their position in the sequence, i.e. stimulus n is associated to stimulus n+1. It has also been shown that spatial and temporal expectations improves the quality of sensory information (Rohenkohl et al., 2012), which might be done by triggering specific neural oscillations to prepare our sensory system to process the temporally and spatially expected information (Cravo et al., 2011). In the present experiment we tested the hypothesis that when learning a predictable sequence of images, expectation mechanisms based on the associative learning between adjacent items in the sequence would induce spontaneous patterns of neural activity selective to the expected event in its absence. Using EEG and classification technique to observe the fine-grain temporal and coarse spatial dimension of neural activity we show here evidence that visual sequence learning induces selective EEG patterns in the absence of an expected stimulus.

We observed that in the absence of the expected event, i.e. Catch trials, the EEG activity in its topographical distribution and in the temporal dynamics at each electrode reflected the expected event. This effect was frequency specific and localized in the high-alpha/low-beta bands. Furthermore the onset of this learning induced selective activity was visible as early as 120ms after the expected event onset, suggesting that this process might be the product of a preparation of the visual system that processed the incoming signal, i.e. a gray square, as if it was the expected stimulus, at least to some extent that yielded these classification accuracies. This latency of peak classification accuracy also suggests that the effects we observed are not merely a voluntary recall of the expected image triggered by the subject but might reflect a spontaneous expectation mechanisms that would be general shared or inherited from spatial navigation and/or associative learning.

These topographical patterns are evoked predominantly in the high-alpha/low-beta frequency bands. This frequency band has been implicated in memory related processes from recent studies on spontaneous context reinstatement and voluntary forgetting (Staudigl et al., 2015; Waldhauser et al., 2015). These studies suggest that the beta-band might underlie the summoning of neural representations, in these cases and in our study visual representations.

This protocol was designed to mimic an "ecological" sequence learning situation in the sense that events occurred in an almost continuous way and subject did not perform any voluntary recall during the trials we used in our analysis, i.e. Catch trials. We naturally and automatically make predictions about our environment based on prior experiences. For example in our daily routines, anticipatory skewing of place fields suggest that we generate predictions when navigating a known environment, e.g. commuting to work, but does not necessitate conscious recall of the next-to-come landmark in the path we take "I'm next to the supermarket so the next thing I should encounter is the gas station". Thus we believe that our study reflects a more passive mode of sequence learning, sharing some similarities with spatial navigation, compared to studies investigating associative learning by requiring subjects to explicitly recall the items' presentation order (Kumaran and Maguire, 2006).

In summary our results show that when learning a predictable sequence of events over and over again, the brain creates expectations that can trigger activation of the expected event neural representations automatically. Expectation mechanisms have been reported to enhance the quality of sensory information (Rohenkohl et al., 2012) in order to prepare our sensory systems to process efficiently and rapidly the expected incoming stimulus and generate predictions that will then be compared with sensory inputs.

The present study also suggests a possible neural mechanism, relying on oscillatory processes, that might underlie the sharpening and reinstatement of neural stimulus template reported in studies on the effect of prior expectations on sensory representations evoked in the introduction of this article (Kok et al., 2012, 2014).

References

Carlson, T., Tovar, D.A., Alink, A., and Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. J. Vis. *13*, 1–1.

Chakravarthi, R., Carlson, T.A., Chaffin, J., Turret, J., and VanRullen, R. (2014). The Temporal Evolution of Coarse Location Coding of Objects: Evidence for Feedback. J. Cogn. Neurosci. *26*, 2370–2384.

Cravo, A.M., Rohenkohl, G., Wyart, V., and Nobre, A.C. (2011). Endogenous modulation of low frequency oscillations by temporal expectations. J. Neurophysiol. *106*, 2964–2972.

Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 815–836.

Fuentemilla, L., Penny, W.D., Cashdollar, N., Bunzeck, N., and Düzel, E. (2010). Theta-Coupled Periodic Replay in Working Memory. Curr. Biol. *20*, 606–612.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al. (2013). MEG and EEG data analysis with MNE-Python. Brain Imaging Methods *7*, 267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hämäläinen, M.S. (2014). MNE software for processing MEG and EEG data. NeuroImage *86*, 446–460.

Isik, L., Meyers, E.M., Leibo, J.Z., and Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. J. Neurophysiol. *111*, 91–102.

Kok, P., Jehee, J.F.M., and de Lange, F.P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron *75*, 265–270.

Kok, P., Failing, M.F., and de Lange, F.P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. J. Cogn. Neurosci. *26*, 1546–1554.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron *60*, 1126–1141.

Kumaran, D., and Maguire, E.A. (2006). The Dynamics of Hippocampal Activation during Encoding of Overlapping Sequences. Neuron *49*, 617–629.

Mayes, A., Montaldi, D., and Migo, E. (2007). Associative memory and the medial temporal lobes. Trends Cogn. Sci. *11*, 126–135.

Mehta, M.R., Barnes, C.A., and McNaughton, B.L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. Proc. Natl. Acad. Sci. *94*, 8918–8921.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature *335*, 817–820.

Moser, E.I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. Annu. Rev. Neurosci. *31*, 69–89.

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. *34*, 171–175.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J Mach Learn Res *12*, 2825–2830.

Reddy, L., Poncet, M., Self, M.W., Peters, J.C., Douw, L., van Dellen, E., Claus, S., Reijneveld, J.C., Baayen, J.C., and Roelfsema, P.R. (2015). Learning of anticipatory responses in single neurons of the human medial temporal lobe. Nat. Commun. *6*, 8556.

Rohenkohl, G., Cravo, A.M., Wyart, V., and Nobre, A.C. (2012). Temporal Expectation Improves the Quality of Sensory Information. J. Neurosci. *32*, 8424–8428.

Schapiro, A.C., Kustner, L.V., and Turk-Browne, N.B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. Curr. Biol. *22*, 1622–1627.

Staudigl, T., Vollmar, C., Noachtar, S., and Hanslmayr, S. (2015). Temporal-Pattern Similarity Analysis Reveals the Beneficial and Detrimental Effects of Context Reinstatement on Human Memory. J. Neurosci. *35*, 5373–5384.

Turk-Browne, N.B., Scholl, B.J., Johnson, M.K., and Chun, M.M. (2010). Implicit Perceptual Anticipation Triggered

by Statistical Learning. J. Neurosci. *30*, 11177–11187.

Waldhauser, G.T., Bäuml, K.-H.T., and Hanslmayr, S. (2015). Brain Oscillations Mediate Successful Suppression of Unwanted Memories. Cereb. Cortex *25*, 4180–4190.

Wallenstein, G.V., Hasselmo, M.E., and Eichenbaum, H. (1998). The hippocampus as an associator of discontiguous events. Trends Neurosci. *21*, 317–323.

C. Discussion

1. Summary of the results

In this first experiment we used EEG recordings during a visual sequence-learning paradigm in order to determine if in a predictable environment the brain generates selective patterns of activity in the absence of an expected event. The results of that experiment suggest that in a context where the environment permits to create expectations about the next-to-come input, e.g. navigation, the brain generates representations that are selective to the expected event in its temporally fixed spatial structure, i.e. scalp distribution, and in its spatially fixed temporal structure, i.e. time course of activity at an electrode. More specifically these spontaneously generated representations appear in the absence of the expected event at a very early latency and in specific frequency bands (high-alpha/low-beta bands). In the spatial domain the informative electrodes were located in centro-frontal, right temporal and occipital sites suggesting a top-down mechanism from a hippocampal and/or frontal system generating the expectation and a posterior sensory system hosting the representations of the expected input.

The task we designed is different from most experiment on sequence learning and memory in that even though subject were instructed to recall the sequence order in Test trials we did not analyze them and focused on Catch trials in which we hypothesized that spontaneous patterns of activity would be selective to the expected event. This emphasizes on the somehow more passive moments of sequence learning was for us a way to study the automatic processes involved in voluntary sequence learning, in contrast with unconscious statistical learning in which the information is not explicitly expressed but resides in the temporal regularities of the environment, for a review see the review chapter on statistical learning by Schapiro and Turk-Browne (2015). Furthermore, subjects were instructed to actively learn the sequence, which almost certainly involved voluntary top-down processes that are absent when subjects are not engaged in the learning task.

2. Spatial navigation and expectation mechanisms

Our study was based in part on results from spatial navigation studies, stemming from the potential sequential learning used to learn about a new environment and results showing anticipatory changes of place fields to signal the next to come position during exploration. But how are these two processes, i.e. predicting future position in an environment and forward prediction in sequence learning, related in terms of neural mechanisms? A possible way of addressing this question would be to test if the amount of selective activity generated by subjects when presented with a Catch trials is correlated to their navigational skills, that have been shown to vary widely across individuals (Wolbers and Hegarty, 2010). If this link is shown to exist it could be a new neural marker of navigational skills and more importantly maybe a marker of early stages of some neurodegenerative diseases, such as Alzheimer, in which some patients progressively lose their spatial navigation faculties (Tangen et al., 2015).

3. Neural representations across cognitive states

a. Cross-classification: classifying Catch events using EEG patterns evoked during perception

In order to understand what kind of representations were evoked during Catch trials we carried out a cross-classification analysis in which we trained classifiers at each time point during Stim trials and tested them at each time point on Catch trials. Performing this analysis permits to compare the neural activity evoked by actually perceiving the expected stimulus to the activity evoked when the subjects were only expecting the stimuli. Therefore it is a way of testing if the neural representations evoked during Catch trials share some similarity with perceptual neural representations, i.e. evoked during Stim trials. As can be seen in Figure 21 the results reach similar classification accuracy as found in ERP classification of Catch trials and in the time-frequency analysis. The largest and highest accuracy cluster can be seen when training the classifier between 600ms and 1000ms and testing it on, i.e. classifying, Catch trials in the 250 to 700ms period. This result

suggests that patterns of activity become similar between image on screen and expectation of image on screen, i.e. Stim and Catch trials, after 600ms of image presentation and for earlier time points in Catch trials. One possible explanation would be that the activity in Stim trials that corresponds to Catch trials is the feedback signal from top-down mechanisms. This would be coherent with the view that in Catch trials the system being left without any visual information to process generates top-down predictions and activates the neural representations of the expected sensory stimulus.

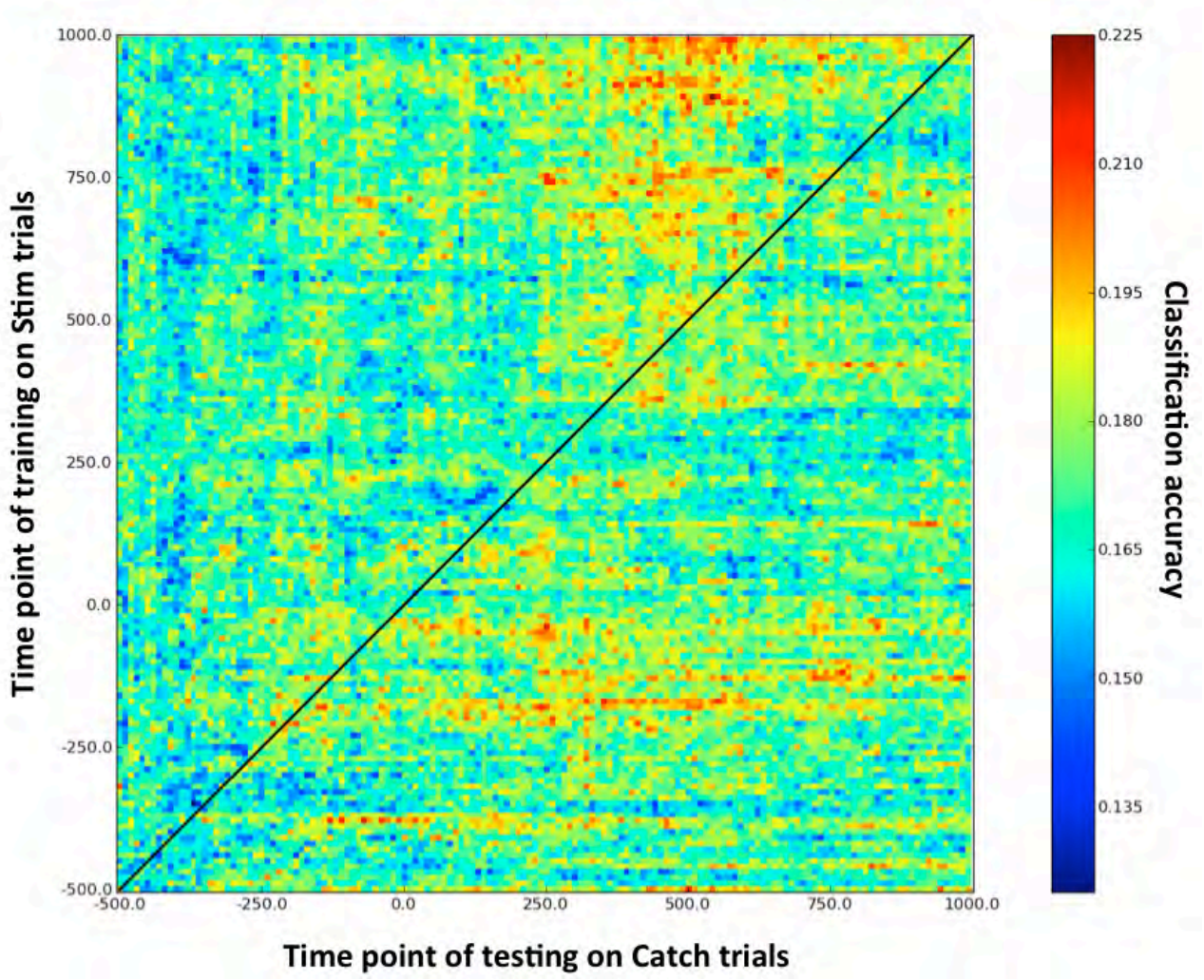

Figure 21. Cross-classification between Stim and Catch trials at each time point of the ERP. Color represents classification accuracy.

b. Mental imagery?

The question of whether perception and mental imagery share the same neural substrates has been the topic of long standing debate (Farah et al., 1988; Kosslyn et al., 2006; Pylyshyn, 2003). Some studies have addressed this issue using neuroimaging methods and asked whether these two cognitive processes have common neural representations. A recent study has tested that hypothesis (Reddy et al., 2010) by cross-classifying, as done in our analysis above, fMRI activity patterns elicited by perceiving certain object categories with patterns elicited while voluntarily creating mental images of these stimuli. This study has shown that while it is possible to classify mental imagery trials alone, it is also possible to cross-classify mental imagery trials using classifiers trained on perception trials. This result is a proof that these two cognitive operations share neural substrates and neural representations.

Even though we cannot say for sure that our cross-classification results presented above are evidence of automatic mental imagery generation, one possible way to resolve this question would be to re-iterate the experiment and conduct mental imagery trials outside of the sequence learning protocol. Doing these mental imagery "localizers" would permit us to then compare:

- If the patterns evoked while imagining the image resemble those evoked when generating selective patterns of activity in the absence of an expected event, i.e. in Catch trials.
- Compare if the time course of classification follows the same dynamics as the one in Catch trials and if the neural substrates underlying the generation of these mental images are common to the ones we observed in Catch trials (informative activity expressed in high-alpha/low-beta bands).

4. Associative learning creating invariance?

Some theories speculate that invariance in the VVP is acquired through associative learning (Földiák, 1991). The idea is that a simple rule could describe how the visual system learns to recognize the same object in different illumination

condition, poses, distances etc. This rule would create invariance by temporal proximity of different views of a same object. A recent study (Isik et al., 2012) have investigated this theory in-silico and they showed that this temporal-association learning rule is accurate by comparing its performance to that of a similar model with hard-wired translation invariance model. The effects we observed in our study might be the first stages of this invariance building process, even though the objects which were associated in our study were not visually similar at all but it asks the question of whether these are universal mechanisms but with different outcomes, e.g. creating invariance or not in our case, or if they are implemented by different mechanisms.

The invariance building by temporal proximity theory surely involves a longer process most certainly requiring consolidation mechanisms. Some studies have investigated at how the MTL encodes associations between stimuli over longer periods, e.g. days, at the neuronal level (Miyashita, 1988; Reddy et al., 2015; Sakai and Miyashita, 1991). For instance Miyashita (1988) observed the formation of neurons that became selective to multiple stimuli when they were repeatedly presented in the same temporal sequence of stimuli. In another study another kind of neuronal selectivity was observed: pair-coding neurons, which either were sensitive to one stimulus and acquired a sensitivity to its paired associate or neurons which were not sensitive to any stimulus and acquired a new selectivity making them respond whether one or the other stimulus of a pair was presented (Sakai and Miyashita, 1991). Even though these studies have shown that it is possible to create new selectivity through long-term associative learning, most of the stimuli used in these experiment were meaningless images, i.e. fractals, leaving aside the question of how do existing stimuli representations, e.g. faces or cars, would be affected by such long-term associative learning. Given that we mostly experience very familiar object categories on a day-to-day basis, i.e. we do not learn new objects everyday, it would be interesting to know how these visual categories are affected by long-term learning of associations between them. We addressed this question in a second experiment using long-term associative learning protocol and fMRI to study the structural flexibility of neural representations in the VVP, i.e. spatial patterns of activity.

Chapter III: The long run, consolidating associations between visual categories

A. Introduction

In the previous study we investigated the effects of learning a sequence of visual events and showed that in the absence of an expected event the evoked EEG topographical patterns are specific to the expected event. This learning setup is considered to be short term learning, subjects were performing the task for 1.5h to 2h. We showed that in this context associative learning creates spontaneous patterns of activity selective to the expected event. There are many reasons to consider short and long term learning differently. For example long-term learning can induce gene expression, protein synthesis and structural changes at the scale of synapses and overall circuitry in the cortex (McGaugh, 2000). Associative learning happening in longer and more repeated exposures, such as training to recognize different races of birds for ornithologists or learning to read, for people who have the chance to be literate, can induce these long-term changes. Would the spontaneous activation we observed due to expectation mechanisms integrate in the long-term structure of visual representations? The concept of long-term storing of information is tightly linked with a memory mechanism we mentioned earlier in the introduction: **system-level consolidation**.

1. Consolidation

When newly acquired memories are stored in long-term memory it is through **consolidation**. This process gradually alleviates the MTL from holding the newly formed connections that constitute this new memory. Evidence of this dissociation first came from neuropsychological studies of MTL lesion or ablation patients who retained remote memories, i.e. acquired a long time before the lesion, but not recent ones (Scoville and Milner, 1957), therefore suggesting that remote memories are stored elsewhere than the lesioned tissue.

A seminal study by Zola-Morgan and Squire (1990) investigated the role of the hippocampus in long-term memory at different delays from the time of acquisition. The rationale of their inquiry comes from the phenomenon of temporally graded retrograde amnesia, where patients suffering from retrograde amnesia following hippocampal damage lost access to recent past more readily than remote past. In

order to test the time-limited role of the hippocampus in memory they had monkeys learn sets of stimulus-reward associations at different delays before their hippocampus was surgically removed. As illustrated on the figure below there are significant differences in retrieval of learned associations at delays of 2 and 4 weeks between lesioned and control monkeys. Performance on the sets learned 8 weeks or more before the lesion was not significantly different whether the monkeys still had a hippocampus or not. This time-limited role of the hippocampus in memory is thought to be the consequence of a gradual development of connections between neocortical sites where associated content is represented, i.e. consolidation.

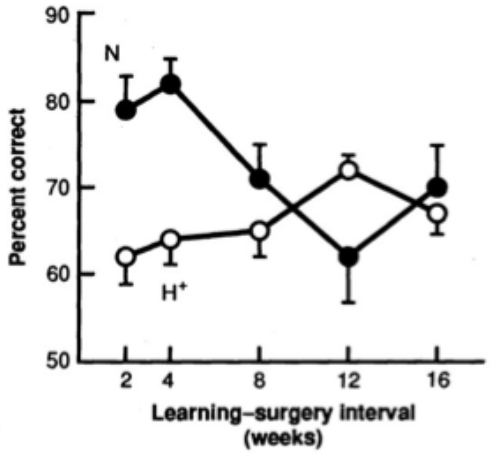

Figure 22 Effects of hippocampal ablation on stimulus-reward association retrieval. This figure shows the percent of correct recall on a pair association task (y-axis) as a function of the delay between pair learning and hippocampal ablation. The filled circles represent control monkeys, i.e. no hippocampal ablation, and the open circles represent monkeys which had underwent hippocampal ablation. (from (Zola-Morgan and Squire, 1990))

These two memory stages, a rapid associative learning between co-occurring information in the hippocampus and a slow, interleaved strengthening of these new links between remote neocortical sites, has been shown to be theoretically necessary when integrating new knowledge in what has been previously learned (McClelland et al., 1995).

Miyashita and colleagues (Miyashita, 1988; Sakai and Miyashita, 1991) have studied neural correlates of associative learning in the MTL and anterior IT. In a series of animal experiments they showed that neurons in the anterior infero-temporal lobe (aIT) could acquire selectivity to pairs of stimuli through associative learning. Monkeys were trained to associate fractal stimuli, i.e. meaningless images, arranged in pairs in a bi-directional manner, i.e. if stimuli 1 and 1' were associated the cue and target could interchangeably be 1 or 1'. Their criterion for acquisition of the associations was 26 out of 30 trials correctly answered two days in a row. After the associations were acquired, responses of some neurons in aIT, near the MTL, were significantly more correlated between paired stimuli than non-paired stimuli showing that some neurons acquired a pair-selectivity code by responding to both pictures of a learned pair.

They then went further in trying to understand how this pair-coding was orchestrated by backward connections from the MTL by lesioning perirhinal (PRC) and enthorinal (ERC) cortices, which provide massive backward ipsilateral projections to aIT (Van Hoesen, 1982; Webster et al., 1991). Using the same task as presented in the previous study and two different sets (set A learned before the lesion and set B after the lesion), they compared the pair-coding acquisition in aIT neurons for the two sets before and after ERC and PRC lesioning with ibotenic acid (i.e. a neurotoxin). They first replicated the pair-coding properties of aIT neurons described in the previous study on set A before the lesion. Following the lesion the monkeys re-learned set A associations to reach a similar performance threshold but they observed a disruption of the pair-coding present before the lesion, i.e. neurons which were previously exhibiting a correlation in spike activity for a pair of stimuli did not show this pair-selectivity after the lesions. The paired associates did not elicit pair-coding properties for set B either, thus showing that ERC and PRC backward projections were necessary to create the pair-coding properties and maintain this functional (re-) organization.

We don't know for how long the learning took place (consolidation?)

Unfortunately the period it took the two monkeys to reach that criterion is not specified in these studies but would be very interesting to better understand how long it took to create pair selectivity in the recorded part of aIT. The acquisition criterion used in Sakai and Miyashita (1991) insures that it was 3 days or more but no further information is given. This information would be very helpful to compare these results with Zola-Morgan & Squire's (1990) on the time-limited role of the hippocampus in memory. For example if consolidation had taken place for the associations learned in Miyashita & al. (1998), such as what is observed for temporally-graded amnesia in MTL patients, removal of backward connections from the MTL should not have disrupted the pair-coding selectivity that appeared for set A. One possibility is that the consolidation process did not take place for long enough to make the newly learned information independent from the hippocampus. In fact it has been shown that the consolidation process will make a memory independent of the hippocampus **and then** the ERC (Izquierdo and Medina, 1997). An interesting question remains: if these learning induced changes in neural responses progress from hippocampus to extra-hippocampus MTL structures, to aIT, can they reach earlier stages of the VVP ?

The line of studies by Miyashita's team provides very insightful information on the way neural selectivity can arise or tune to new stimuli from associative learning, yet it is still unknown how these results connect with consolidation processes from Zola-Morgan and Squire's study (1990) and the time-course of this new selectivity acquisition by neurons of the neocortex.

Functional Magnetic Resonance Imaging

In contrast with electrophysiology, fMRI offers the possibility to record neural activity from the whole brain or at least at a greater spatial scale and thus try to understand its functional organization from a larger point of view.

- Discovery of BOLD signal by Seiji Ogawa (1990)

The use of magnetic resonance imaging to record neural activity is based on the fact that **deoxygenated** hemoglobin and **oxygenated** hemoglobin have different magnetic properties and that **cerebral blood flow** and **neural activity** are

coupled. The first study introducing the use of MRI machines to study neural activity was published by Ogawa & al. (1990), in which they described the **B**lood **O**xygenated **L**evel-**D**ependent (**BOLD**) signal, which is sensitive to blood flow and the ratio of oxygenated versus deoxygenated blood cells. Since then it has become the most popular tool to study large-scale brain activity and test spatially relevant research hypotheses.

Even though we still don't fully understand the mechanisms underlying BOLD signal and physiological markers of neuronal activity such as Local Field Potentials (LFP), Multi-Unit Activity (MUA) or action potentials, a great number of studies have investigated this link. As shown in Figure 23 the fMRI BOLD signal is the result of a process influenced by many factors. In a pioneering study Logothetis & al (2001) were able to show that LFP yields the best approximation of the BOLD signal response, thus fMRI seems to reflect input and intracortical processing of the recorded area rather than its spiking activity.

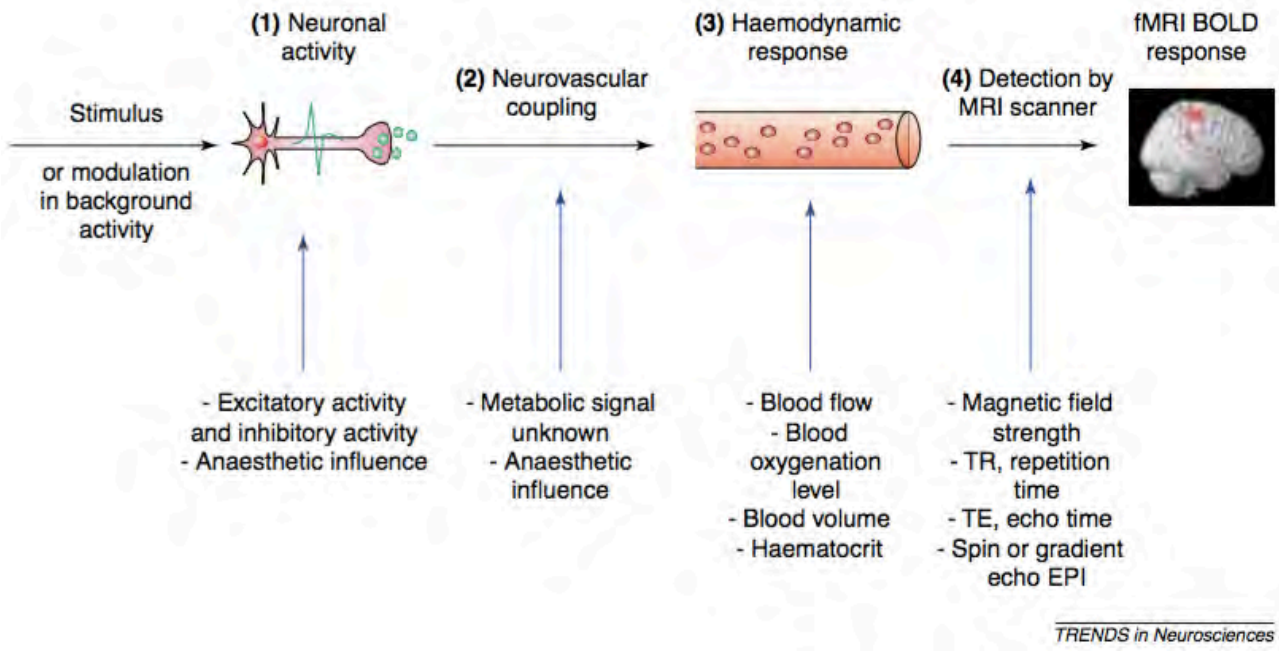

Figure 23. Factors influencing the BOLD signal (haematocrit = volume percentage (%) of red blood cells in blood) (from (Arthurs and Boniface, 2002))

Long-term memory and consolidation

Thanks to fMRI's ability to record whole brain activity, the fact that it is non-invasive and can be used longitudinally, Takashima & al (2006) tackled the question of which neural structures are involved in memory retrieval at different stages of consolidation with precise temporal intervals: days, weeks and months.

A pending question in research on declarative memory consolidation at that time was whether the hippocampus participates in remote memory retrieval by conserving the nodes of the contents forming a memory or if other structures take on this putative role in long-term memory. Furthermore until recently the time-course of long-term memory storage in humans was poorly described, ranging from decades for patients with MTL lesions who retained some hippocampus-independent memories to a few weeks in animal studies, but these latter evidences still needed to be proven to be equivalent in the human neural system and that it corresponds behaviorally to patient studies. Takashima & al (2006) addressed these questions by conducting an fMRI activation study on long-term memory of visual stimuli of natural landscapes. Subjects first underwent a learning paradigm with feedback to memorize the 320 images, which constituted the remote set, and subsequently participated in four fMRI sessions: on day 1, 2, 30 and 90 (day 1 being the day of the remote set learning session). At each session subjects had to learn a new set of stimuli outside the MRI, i.e. the recent set. Inside the MRI they were presented with a subset of stimuli from the remote and recent sets and had to indicate if they saw them before in order to compare neural activity when correctly recognizing stimuli acquired remotely (at increasing time-intervals) to recently. Animal studies have shown that lesions to the Ventro-Medial Pre-Frontal cortex (VMPFC) led to impairment in long-term declarative memories, showing a putative role of the VMPFC in keeping the "nodes" after consolidation, hence not anymore in the hippocampus. Their hypothesis was that as the time interval between first exposition and recognition increases there would be a gradually decreased activity of the hippocampus along with an increased activity of VMPFC.

Using classical fMRI activity contrast methods between correct recognition of stimuli from the remote and recent sets they observed progressively decreasing hippocampal activity and increasing VMPFC activity as the delay between first exposition to the stimulus and correct recognition increased, from 1 to 90 days before the scan, see Figure 24.

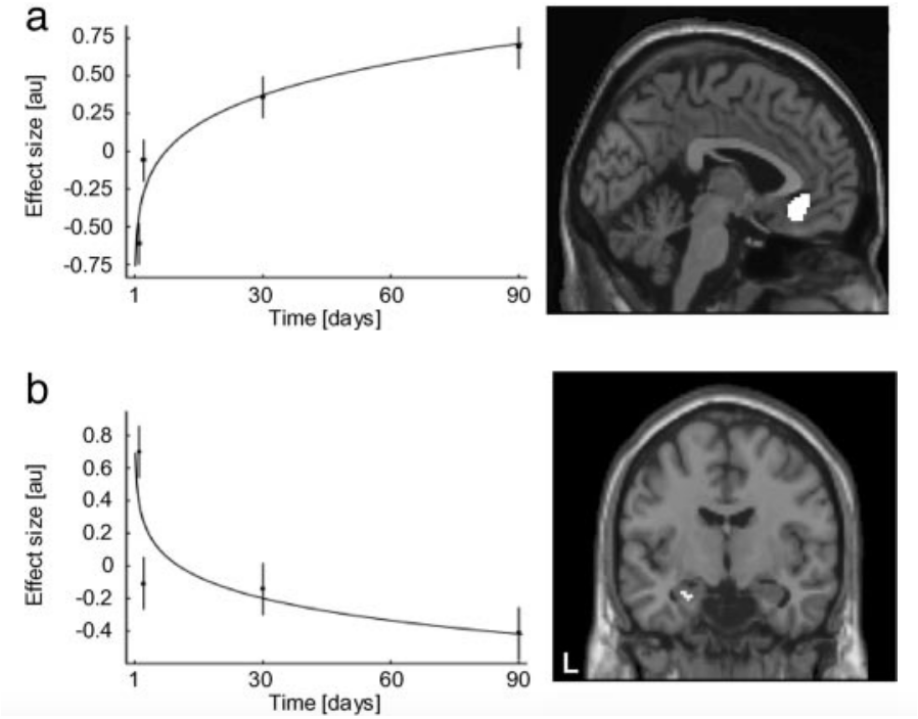

Figure 24 Hippocampal and VMPFC effects of consolidation. (a) This figure depicts the increasing activity in VMPFC as a function of the delay between the first exposition to the remote set and the recall. (b) This figure shows the decreasing activity in the hippocampus as a function of delay between the exposition to the remote set and recall. From (Takashima et al., 2006)

In agreement with their hypotheses the hippocampus disengaged from the retrieval process with time and the VMPFC had the opposite effect, being more activated for retrieval as time passed by. Using fMRI they were thus able to study how consolidation affects the involvement of different areas in retrieval, but what about the neural representations specific to the content of retrieved memories? Are they subject to changes due to learning and consolidation? Some theories argue that there should be a strengthening of connections between neural assemblies that have been associated to form the memory by consolidation

(Frankland and Bontempi, 2005). For example we could ask if the VVP held visual representations of the remembered scenes from the remote set, undergo any modification to their structure due to consolidation?

Can object representations in the adult VVP be shaped by experience?

As was discussed in the introduction, development of the VVP takes years to reach adult-like anatomical and functional structure, as do the perceptual abilities it implements (Grill-Spector et al., 2008; de Haan et al., 2002; Mondloch et al., 2003). But how flexible is this functional organization in adults?

Investigation of the neural basis of perceptual learning and more specifically visual object learning has shown that VVP representations can be modulated through experience. In a recent review on "The neural basis of visual object learning" by Op de Beeck & Baker (2010) it is argued that training as a discrimination or categorization task will induce changes at the neuronal level or in neuronal populations depending on the usefulness of their pre-learning response properties and whether it suffices or not to perform the training task. For neurons the properties can be summarized as the optimal stimulus (the one that drives the largest response of the neuron or the population), the selectivity (the specificity of the neuron's response in the stimulus space). For neuronal populations other properties are of interest such as the sparseness, only a few neurons were impacted, and clustering, how clustered were the impacted neurons, of the changes in population response. fMRI can be considered as a neuronal population recording technique, although situated at the extreme end of the spectrum from electrophysiological recordings of single neuron to populations of neurons. A study by Op de Beeck & al (2006) investigated the changes induced by discrimination training on new object categories and showed that this training increased the selectivity to the trained category but also changed the pattern of selectivity. This result shows that the object representations in the VVP can be modulated by experience in adult humans. But as other results cited in this article investigating flexibility of object representations in the VVP was almost exclusively done on **novel objects**, this leaves open the question of **are existing neural representations of visual categories still flexible?** More specifically we investigated how multi-voxel patterns representations were modulated by associative learning between existing visual object categories.

B. Paper

Multi-voxel Object Representations in Adult Human Visual Cortex are Flexible: an Associative Learning study.

Mehdi Senoussi [1, 2], Isabelle Berry [3, 4], Rufin VanRullen [1, 2], Leila Reddy [1, 2,*]

1 Université de Toulouse ; UPS ; Centre de Recherche Cerveau et Cognition ; 31052 Toulouse, France 2 CNRS ; CerCo ; 31052 Toulouse France 3 Inserm Imagerie cérébrale et handicaps neurologiques UMR 825 F-31059 Toulouse, France 4 Centre Hospitalier Universitaire de Toulouse Pôle Neurosciences CHU Purpan, Place du Dr Baylac, F-31059 Toulouse Cedex 9, France

Multivoxel Object Representations in Adult Human Visual Cortex Are Flexible: An Associative Learning Study

Mehdi Senoussi[1,2], Isabelle Berry[3,4], Rufin VanRullen[1,2], and Leila Reddy[1,2]

Abstract

■ Learning associations between co-occurring events enables us to extract structure from our environment. Medial-temporal lobe structures are critical for associative learning. However, the role of the ventral visual pathway (VVP) in associative learning is not clear. Do multivoxel object representations in the VVP reflect newly formed associations? We show that VVP multivoxel representations become more similar to each other after human participants learn arbitrary new associations between pairs of unrelated objects (faces, houses, cars, chairs). Participants were scanned before and after 15 days of associative learning. To evaluate how object representations changed, a classifier was trained on discriminating two nonassociated categories (e.g., faces/houses) and tested on discriminating their paired associates (e.g., cars/chairs). Because the associations were arbitrary and counterbalanced across participants, there was initially no particular reason for this cross-classification decision to tend toward either alternative. Nonetheless, after learning, cross-classification performance increased in the VVP (but not hippocampus), on average by 3.3%, with some voxels showing increases of up to 10%. For example, a chair multivoxel representation that initially resembled neither face nor house representations was, after learning, classified as more similar to that of faces for participants who associated chairs with faces and to that of houses for participants who associated chairs with houses. Additionally, learning produced long-lasting perceptual consequences. In a behavioral priming experiment performed several months later, the change in cross-classification performance was correlated with the degree of priming. Thus, VVP multivoxel representations are not static but become more similar to each other after associative learning. ■

INTRODUCTION

We can rapidly and accurately detect and categorize objects even when they are flashed for just a fraction of a second. This astonishing ability relies on the ventral visual pathway (VVP), a neural system that extends from the occipital cortex to lateral and ventral regions of the temporal lobe (Grill-Spector, 2003). The VVP is not organized in a homogenous fashion (Grill-Spector & Malach, 2004). Instead, this expanse of cortex is dotted with several smaller regions that respond preferentially to specific classes of stimuli (e.g., faces, places, objects, or bodies; Downing, Jiang, Shuman, & Kanwisher, 2001; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). This underlying architecture is remarkably consistent across normal, healthy participants (Haxby et al., 2011).

Object category representations in the VVP can be described at two different levels: in the activity of large-scale multivoxel patterns (MVPs) or at the level of the object selectivity of individual neurons (Reddy & Kanwisher, 2006). Although it is difficult to measure the selectivity of single neurons in the human brain, it is now well established that object category information is also reflected in the large-scale MVPs of activity that can be recorded with fMRI. Indeed, decoding studies have shown that category information is explicit in these response patterns (Op de Beeck, Brants, Baeck, & Wagemans, 2010; Reddy & Kanwisher, 2007; Spiridon & Kanwisher, 2002; Haxby et al., 2001). Here we ask if MVPs for well-learned categories still maintain flexibility related to visual experience in the adult brain.

Specifically, in this study, we directly test if large-scale representations for highly familiar categories in the VVP become more similar to each other when pairs of categories are behaviorally associated through extensive training. At the neuronal level, anterior ventral temporal cortex and medial-temporal lobe (MTL) structures have been implicated in associative learning in both monkeys (Wirth et al., 2003; Messinger, Squire, Zola, & Albright, 2001; Miyashita & Chang, 1988) and humans (Ison, Quian Quiroga, & Fried, 2015; Reddy et al., 2015). However, here we show that preexisting multivoxel representations for familiar objects (faces, houses, chairs, cars) in ventral visual cortex shift in a concerted way in a high-dimensional multivoxel space once two categories become perceptually related.

We scanned 20 observers before and after they learned arbitrary associations between different object categories (faces, houses, cars, chairs) and investigated changes in the large-scale category representations with MVP analysis

[1]Université de Toulouse, [2]CNRS, CerCo, Toulouse, France, [3]Inserm Imagerie cérébrale et handicaps neurologiques UMR 825, Toulouse, France, [4]Centre Hospitalier Universitaire de Toulouse Pôle Neurosciences CHU Purpan

Journal of Cognitive Neuroscience 28:6, pp. 852–868
doi:10.1162/jocn_a_00933

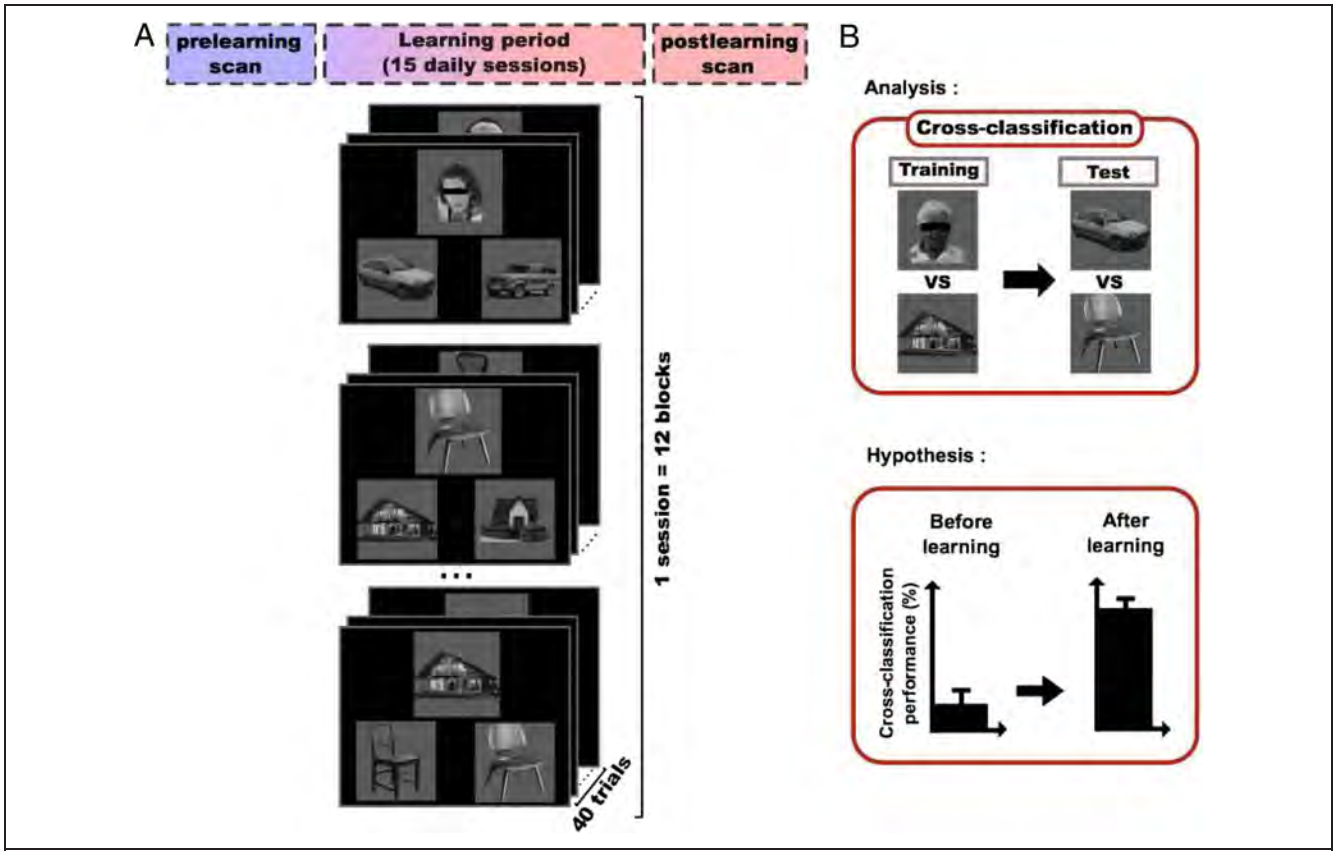

Figure 1. Experimental protocol and hypothesis. (A) Each participant followed a three-step procedure. In the first step, participants performed a prelearning scan in which they viewed blocks of faces, houses, chairs, cars, and scrambled images. Next, in 15 daily sessions, participants performed a learning task in which they learned arbitrary associations between members of the different categories. In this example, faces are paired with cars and houses with chairs. Category pairings were counterbalanced across subjects. Each learning session consisted of 12 blocks of 40 trials. On each trial, participants were presented with a main stimulus (e.g., a face) and two choice stimuli from the associated category (e.g., two cars) and had to decide which of the choice stimuli was paired with the main stimulus (by pressing the left or right arrow keys on the keyboard). After the learning sessions, participants performed a postlearning scan that was identical to the prelearning scan except that the block order was randomized. (B) To evaluate the similarity between category representations before and after learning we used a cross-classification procedure with the searchlight method. An SVM classifier was trained to distinguish between two categories and tested on their associated categories. We hypothesized that after learning, we would see an increase in cross-classification performance suggesting that the multivoxel representations of the paired categories had become more similar to each other.

methods. In particular, we trained a support vector machine (SVM) classifier to discriminate between two nonassociated object categories (e.g., houses vs. faces) and then tested it on discriminating between their paired associates (e.g., cars vs. chairs). We hypothesized that after learning we would see an increase in this cross-classification performance. Because a classification decision reflects the distance and the relative position of test patterns in a multidimensional space, an increase in cross-classification performance after learning would imply that the representations of the paired categories had moved in a high-dimensional multivoxel space or, equivalently, had become more similar to each other.

Using this cross-classification approach, we found an increase in decoding performance after learning, suggesting that large-scale fMRI response patterns in the VVP for associated object categories become more similar to each other. In other words, in an example participant who associated faces with chairs and houses with cars, face MVPs became more similar to chair MVPs and house MVPs became more similar to car MVPs after learning. This shift in category representations had perceptual consequences, as measured by a behavioral priming task performed several months after the associations had been learned. Specifically, we found that a given category facilitated the processing of its paired associate relative to the processing of a nonassociated category. In addition, this priming effect was correlated across participants with the overall amount by which the category representations shifted as a result of learning.

METHODS

Participants and Stimuli

Twenty-one participants were recruited for this study (10 women, mean age = 24 years, range = 19–35 years). One participant was excluded from the study because of excessive motion in the scanner. All participants had normal or corrected-to-normal vision and reported no

history of neurological problems. All participants provided written informed consent and received monetary compensation for their participation. The local ethics committee for human experimentation approved all procedures.

Ten stimuli from each of four categories (faces, houses, chairs, cars) were gathered from different sources on the Internet. These images were then transformed to gray-scale and pasted on a 500 × 500 pixels gray canvas. To avoid low-level category confounds, we normalized categories in luminance, contrast, and size. We then generated a scrambled version of each image for the functional ROI localizers.

Experimental Protocol

The experimental protocol consisted of three phases: a prelearning fMRI scan, an associative learning task outside the scanner over 15 days, and a postlearning fMRI scan.

During the fMRI scans, stimuli were presented with the VisionEgg toolbox (Straw, 2008). Each fMRI run consisted of four blocks each of the four categories (faces, houses, cars, and chairs) and scrambled images and five blocks of fixation. Each block was 16 sec long. The fixation blocks occurred after every five visual stimulation blocks. In each visual stimulation block, 16 stimuli were presented, each for 800 msec followed by an ISI of 200 msec. Participants were instructed to press a button when the same image was presented on two successive trials (1-back task). Each fMRI session consisted of eight runs that lasted approximately 6 min and 45 sec each. The pre- and postlearning fMRI sessions were identical, except for the block and stimulus order, which were randomized in each run.

In between the fMRI sessions, participants underwent 15 daily learning sessions during which they learned associations between exemplars of the object categories (e.g., each face was associated with a given car, and each house with a given chair). Each 20-min session consisted of 12 blocks of 40 trials. Each trial lasted up to 3 sec with an intertrial interval of 0.750 sec. On each trial, participants were presented with a main stimulus (e.g., a chair) and two choice stimuli (e.g., two houses) and had to decide (by pressing one of two keys on the keyboard) which of the choice stimuli was the correct associate of the main stimulus (Figure 1A). Exemplars of each category served as the main stimulus or choice stimuli on different blocks. Ten exemplars per category were used. Learning was achieved by trial and error, and negative auditory feedback was provided on incorrect trials. The category pairings were counterbalanced across participants: Half of the participants associated faces with cars and houses with chairs, and the other half associated faces with chairs and houses with cars.

Priming Experiment

The priming experiment was performed on average 14.1 months after the postlearning fMRI scan on 14 of the original 20 participants. Before participants performed the priming experiment, they underwent three training sessions on the main associative learning paradigm. They then performed two sessions of the priming experiment on two days.

To avoid low-level priming effects, we equalized all stimuli in the Fourier amplitude spectrum. On each trial of the priming experiment, participants were presented with a prime stimulus for 100 msec followed by a target stimulus for 2 sec and instructed to report the category of the target stimulus (Figure 2). The intertrial interval was 1000 msec, with a jitter of 500, 750, or 1000 msec. After each trial, the fixation cross turned to a dash for 1 sec and turned back to a cross to signal the beginning of the next trial. The prime and target stimuli were exemplars of the four object categories (faces, houses, chairs, cars). Within a block of trials, only two categories were targets (e.g., cars and chairs in blocks when participants were asked to

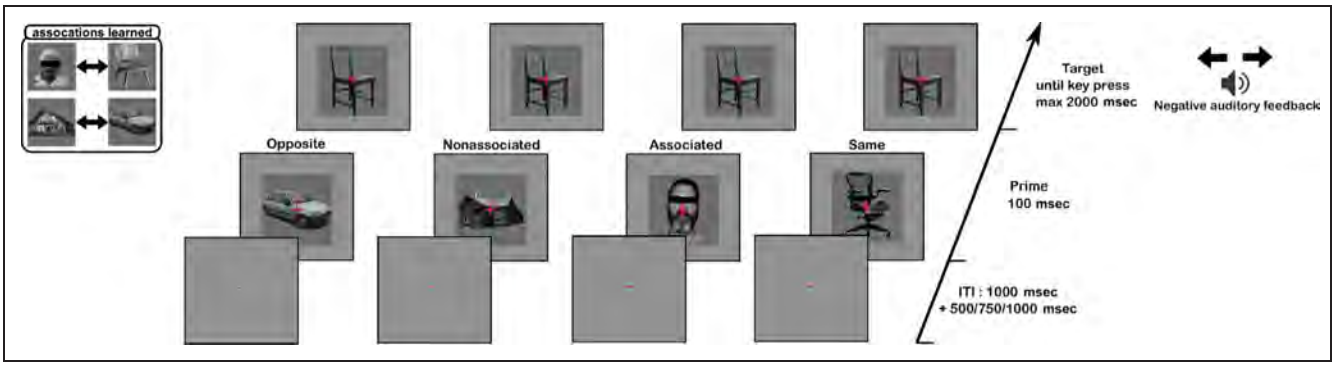

Figure 2. Behavioral priming task experiment design: On each trial of the priming experiment, participants were presented with a prime stimulus followed by a target stimulus and instructed to report the category of the target stimulus. The prime and target stimuli were exemplars of the four object categories (faces, houses, chairs, cars). There were four types of trials: When the primes and targets were different exemplars from the same category ("same" trials), when the prime and target were from opposite categories with respect to the category discrimination task ("opposite" trials), and when the prime and target were from associated/nonassociated categories.

discriminate cars from chairs), whereas all four categories could serve as primes. There were four types of trials: when the prime and target category matched ("same" trials), when the prime and target were from opposite categories with respect to the category discrimination task (e.g., the prime was a car and the target was a chair in a block when participants were instructed to discriminate cars from chairs, "opposite" trials), and when the prime and target were from associated/nonassociated categories (e.g., faces/houses, "associated"/"nonassociated" trials). For "associated" trials, the primes and targets were the pairs learned during the associative learning paradigm, for example, a participant who had learned to associate face1 with chair5 was presented with face1 as a prime when chair5 was the target on "associated" trials, in a block where participants were instructed to discriminate cars from chairs. Participants were instructed to respond as fast as possible on each trial. Each participant performed eight blocks of 250 trials. Trials were randomized within each block. Participants performed the priming experiment over 2 days. On the first day, the targets were cars and chairs, each with their own response button (left and right, respectively). On the second day, the targets were faces and houses, each with their own response button (up and down, respectively). We chose this design to avoid confusing participants by switching instructions within a single experiment session.

fMRI Data Acquisition and Analysis

fMRI data were collected on a 3T Philips (Amsterdam, The Netherlands) ACHIEVA scanner (gradient-echo pulse sequence, repetition time = 2 sec, echo time [TE] = 35 msec, 30 slices with a 32-channel head coil, slice thickness = 2 mm, in-plane voxel dimensions 1.88 × 1.88 mm). The slices were positioned to cover the entire temporal and occipital lobes. High-resolution anatomical images were also acquired per participant (1 × 1 × 1 mm voxels, repetition time = 8.13 msec, TE = 3.74 msec, 170 sagittal slices). Data analysis was performed with FreeSurfer and the FreeSurfer Functional Analysis Stream (FS-FAST; surfer.nmr.mgh.harvard.edu), custom Matlab scripts, and the PyMVPA toolbox (www.pymvpa.org/; Hanke et al., 2009). Similar results were also obtained with the SearchMight Toolbox (www.princeton.edu/~fpereira/searchmight/).

Preprocessing followed the FS-FAST processing stream. All images were motion-corrected (using AFNI with standard parameters), slice time-corrected, intensity-normalized, and smoothed with a 3-mm FWHM Gaussian kernel. We then estimated the beta weights using a general linear model (GLM) for the five stimulus conditions (faces, houses, cars, chairs, and scrambled) in each participant. The betas were computed on whole-run data. There were eight runs in each scan session and four blocks of 16 sec of each condition in each run. We obtained eight beta images per condition (i.e., one from each run) from each scanning session from the FS-FAST processing stream. The GLM fit the hemodynamic response with a gamma function (delta = 2.25, tau = 1.25) and modeled the drift with an order 1 polynomial. For all other parameters of the GLM, we used the default settings from FS-FAST. Finally, the beta-weight volumes were normalized on the MNI305 brain, and we used these volumes as inputs for the searchlight analysis. Similar results were obtained when the searchlight analysis was performed in the native space of each participant.

ROIs

ROIs were defined manually in each participant's native space using an independent analysis. Fusiform face area (FFA) was defined as the set of contiguous voxels in the fusiform gyrus that exhibited greater activation for faces than houses ($p < 10^{-5}$, uncorrected). Parahippocampal place area (PPA) was defined as the set of contiguous voxels in the parahippocampal gyrus that exhibited greater activation for houses than faces ($p < 10^{-5}$, uncorrected). lateral occipital complex (LOC) was defined as the set of voxels in the inferior occipital and temporal cortices that exhibited greater activation for cars and chairs than scrambled images ($p < 10^{-5}$, uncorrected). The anterior and posterior subdivisions of LOC (lateral occipital [LO] and posterior fusiform [pF]) were also identified for each participant. The hippocampus, V1, and V2 were defined using anatomical landmarks for each participant in FreeSurfer. The average ROIs displayed in Figure 7 were computed by selecting voxels that were common to at least 60% of the ROIs defined in individual participants. Note that the ROI analyses were performed in each participant's individual ROIs, and the average ROI is used for display purposes only.

Multivariate Analysis

The searchlight analysis was performed with the *CrossValidation*, *HalfPartitioner*, *LinearCSVMC*, and *sphere_searchlight* functions of the PyMVPA toolbox using default settings. A linear SVM with default settings from the PyMVPA toolbox was used to perform a cross-classification analysis within each searchlight. We used searchlights of different radii (1–14 voxels) that we moved along the MNI305 volumes. For each participant, within each searchlight, an SVM classifier was trained on the fMRI patterns for two nonassociated categories for that participant (e.g., faces vs. houses) and tested on the corresponding associated categories (e.g., cars vs. chairs). Additionally, the symmetric classification was also performed (i.e., in the example here, a car–chair classifier was tested on a face–house discrimination). The average of the two classification scores was reported as the cross-classification performance for the voxel at the center of the searchlight. The input to the classifiers were eight MVPs for each condition. For example, when training a classifier on a face versus house discrimination and testing it on a car versus chair discrimination, the

classifier was trained on eight MVPs of face betas and eight MVPs of house betas and tested on eight MVPs of car betas and eight MVPs of chair betas.

The same stimuli and data sets were used for the experimental sessions and for defining the functional ROIs. However, our analysis is free of the double-dipping problem (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009) because orthogonal contrasts were used in defining the ROIs versus in the cross-classification analysis. For instance, when defining the FFA, we used a faces–houses contrast. On the other hand, the cross-classification analysis tested a face/house classifier on a car/chair discrimination. Defining our FFA with a face–house contrast guarantees that a face/house classifier in these voxels would perform superbly on a face/house discrimination of the same data (i.e., a circular analysis). However, there is no reason for performance of the face/house classifier on a car/chair discrimination task to benefit from this method of voxel selection.

Correct or incorrect classification depended on the association learned by the particular participant. For example, for participants who learned face–car and house–chair associations, cross-classification would be deemed as correct if the face–house classifier classified the MVP elicited by cars as faces and the MVP elicited by chairs as houses. This procedure produces cross-classification performances ranging from 0 to 1. A performance of 1 means that the classifier always considered patterns of associated categories as being more similar, a performance of 0 means that it always considered patterns of nonassociated categories as being more similar, and a score of 0.5 means that the classifier did not have any bias between the categories. This procedure could be done in two ways, because there were two pairs of associations: training the classifier on faces and houses and testing it on cars and chairs, or training it on cars and chairs and testing it on faces and houses. The results of these two analyses were equivalent so the final cross-classification performance values were averaged across the two analyses.

Note that the cross-classification approach might be a more sensitive test of learning-induced flexibility than a direct classification test on the associated category pairs because a priori, a chair pattern should fall roughly halfway between a face and a house pattern (i.e., 50% classification performance), so a small shift of the chair pattern toward the face pattern could result in a sizeable change in cross-classification performance. On the other hand, if face and chair patterns become more similar in a multidimensional space, they might still be far enough apart that a direct face/chair classifier would never confuse a chair with a face and thus learning would not seem to modify classification accuracy.

The searchlight analysis was performed across the entire scanned functional volume as well as in the specific

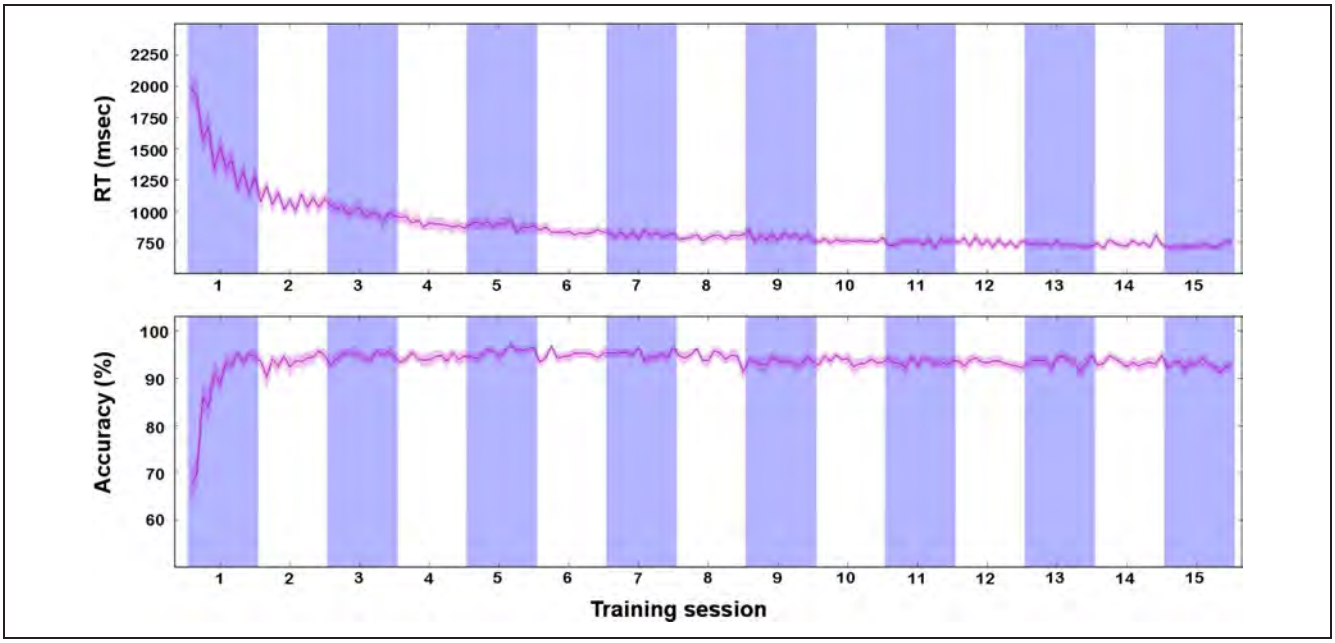

Figure 3. Behavioral results during learning. Each participant performed 15 learning sessions outside the scanner. The RTs and accuracies in each session are shown here (individual sessions indicated by the blue and white areas). RTs (top plot) decreased steadily (one-way, random-effects ANOVA on log(RT): $F(14, 266) = 128.61, p < 10^{-6}$), and stabilized after the tenth session (post hoc Tukey's HSD ($p < .05$)). For statistical tests only (but not for display purposes), the RTs were log-transformed to satisfy the constraints of normality. Accuracy (bottom plot) was computed for each session as the proportion of trials where the participant responded correctly. Accuracy on the learning task was at or above 90% by the end of the first learning session for most participants (19 of 20) and then stabilized by the second session (one-way, random-effects ANOVA: $F(14, 266) = 9.38, p < 10^{-6}$, post hoc Tukey's HSD ($p < .05$)). The pink lines correspond to the average across participants of all trials of each block in each session, and the shaded area is the *SEM*.

ROIs defined for each participant. For the ROI-specific analyses, we used a fixed-size searchlight of radius 3 voxels (i.e., a searchlight consisting of 123 voxels).

Statistical Analysis

The statistical significance of the difference between the pre- and postlearning distributions was evaluated using two-tailed one-sample t tests across participants. To assess the statistical significance of the voxels that showed the largest cross-classification shifts in Figure 4 (and in the corresponding surface maps in Figure 7), we used a nonparametric test in which we shuffled the labels of the pre- and postlearning sessions for each voxel and for each participant independently to simulate the null hypothesis that there was no difference between these sessions for each voxel. The surrogate distributions were computed 2000 times per participant. The p value of each voxel was assigned by comparing this voxel's cross-classification shift to the corresponding surrogate values (i.e., 2000 iterations × 78,842 voxels).

Correlation Analysis

In the pF, FFA, PPA, and LO, a searchlight of radius 3 voxels, centered on each voxel, was trained and tested on discriminating the four categories (faces, houses, chairs, cars) from each other prior to learning. This four-way classification analysis was performed individually for each participant in the MNI305 space and then averaged across participants to obtain an average four-way classification performance value for each voxel. This average four-way classification performance value was then correlated with the change in cross-classification performance of each voxel (also averaged across participants to obtain one performance value per voxel). The parameters of the four-way classifier were identical to the cross-classification classifier (see above). The four-way classifier was trained on blocks of data from all runs but one and tested on the remaining run (leave-one-run-out cross-validation). This correlation analysis was performed on the average ROIs computed by selecting voxels that were common to at least 60% of the ROIs defined in individual participants.

To make the correlations comparable across ROIs, we equalized the number of voxels in each ROI before computing the correlation value. Specifically, we resampled each ROI 100,000 times, each time randomly choosing 162 voxels (that corresponded to the size of the smallest ROI, pF) and computing the Pearson's r value in each resample. The reported r^2 values correspond to the square of the average r values of these resamples.

RESULTS

Twenty observers were scanned before and after they learned arbitrary associations between pairs of different object categories (Figure 1A). During these pre- and postlearning fMRI scans, the participants viewed 10 exemplars each of faces, houses, chairs, cars, and scrambled images in different blocks. Participants performed a 1-back task, in which they responded if the same image had been presented on two successive trials. Note that, in the scanner, the image presentation order and the 1-back behavioral task were independent of the associations learned by the participants outside the scanner. These scans simply allowed us to obtain pre- and postlearning MVPs for the four object categories.

In between these two scan sessions, participants learned arbitrary associations between different exemplars of the four object categories (e.g., each face was associated with a car/each house with a chair; Figure 1A). Most participants achieved greater than 90% accuracy by the end of the first session and continued to improve until behavioral measures of learning stabilized by the tenth session. Participants continued to train even after performance had stabilized (Figure 3).

As mentioned above, we trained an SVM classifier to discriminate between two nonassociated categories (e.g., faces and houses) and tested it on discriminating their paired associates (e.g., cars vs. chairs). We hypothesized that after learning we would see an increase in this cross-classification performance (Figure 1B), suggesting that the multivoxel representations of the paired categories had become more similar to each other. Because we had no strong a priori expectation about where these learning-related changes might occur, we used the searchlight method to explore different areas of the VVP (Kriegeskorte, Goebel, & Bandettini, 2006).

To perform the cross-classification procedure, we realigned each participant's functional volume to the MNI305 brain to make comparisons across participants. We moved a spherical searchlight along each participant's realigned functional volume and, at each voxel, calculated the cross-classification performance from the MVPs falling within the searchlight centered on that voxel. More specifically, for an example participant who had learned to associate faces with cars and houses with chairs, we tested the performance of a face–house classifier on car–chair discrimination and a car–chair classifier on face–house discrimination within the searchlight. Note that there is no "correct" answer for either of these classifiers, as the associations were arbitrarily determined—we simply assumed that, faced with a meaningless choice, the classifier would tend to choose the label of the associated category. The average of these two classification scores was the cross-classification score attributed to the voxel at the searchlight center. We performed this analysis separately on the MVPs from the pre- and postlearning scans and evaluated how cross-classification performance changed after learning.

The pre- and postlearning distributions of cross-classification performance across all the voxels in the scanned volume were averaged across the 20 participants and are shown in Figure 4A. To obtain optimal classification

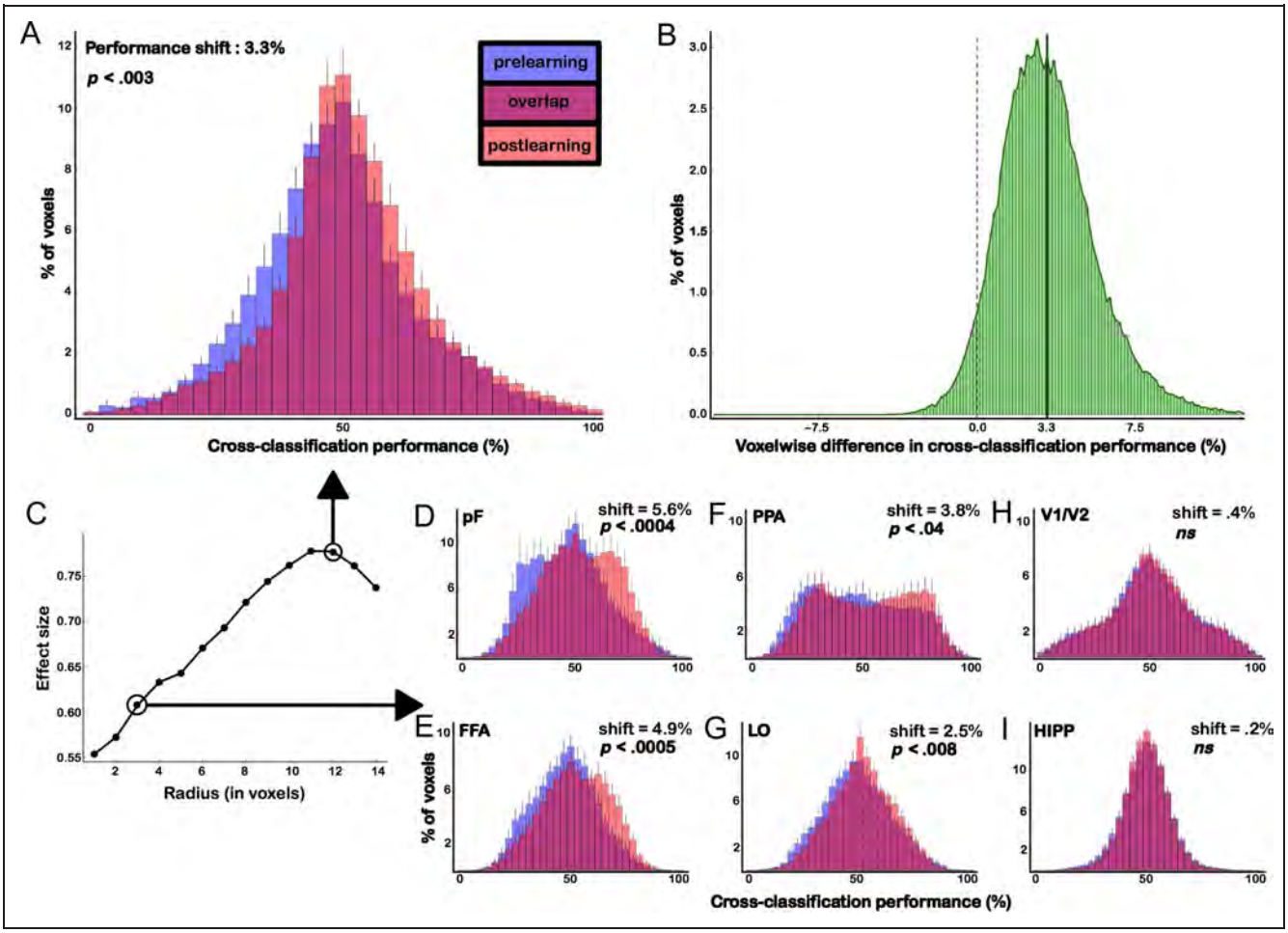

Figure 4. Cross-classification performance before and after learning. (A) Histograms showing the distribution of cross-classification performance across all voxels in the scanned volume, obtained with a searchlight of radius 12 voxels, averaged across 20 participants. Prelearning performance values are in blue, and postlearning performance values are in pink. The overlap between the pre- and postlearning distributions is shown in purple. Across all voxels, there was a significant increase in performance after learning (3.3 ± 0.95%; ($t(19) = 3.39$, $p < .005$). (B) Distribution of the voxelwise difference between the pre- and postlearning performance values. As expected from A, the average voxelwise difference was 3.3%. However, although the shift was absent or only moderate for some voxels, a number of voxels shifted by more than 10% on average. (C) The effect size of the difference between pre- and postlearning cross-classification distributions obtained with searchlights of different radii. (D–G) Same as in A for different ROIs, obtained with a searchlight of radius 3 voxels. The significance and shift of the difference between the pre- and postlearning distributions are indicated for each panel.

performance, the size of the searchlight must be commensurate with the size of the region where the effects occur (Kriegeskorte et al., 2006). Accordingly, we tested the effect of varying the searchlight radius on cross-classification performance. In the whole-brain analysis we obtained optimal cross-classification performance with a searchlight of radius 12 voxels (Figure 4A), but similarly significant effects were also obtained with searchlights of other radii from 1 to 14 voxels (Figure 4C). The pre- and postlearning average cross-classification performance values were 48.3% and 51.7%, respectively, and not significantly different from chance levels (50%, $t(19) = 1.55$, $p = .13$ for the prelearning distribution and $t(19) = 1.82$, $p = .08$ for the postlearning distribution). However, between the two learning sessions, we observed a significant increase of 3.3 ± 0.95% (mean ± *SEM*) in the average cross-classification performance over all voxels in the scanned volume. We statistically evaluated this difference in mean cross-classification performance between the two scan sessions using a two-tailed, paired t test of average pre- and postlearning performances (with each participant contributing one global cross-classification performance value to the statistical test, thus avoiding multiple comparisons across voxels or brain regions and warranting the assumption of independence between measurements ($t(19) = 3.39$, $p < .003$). Furthermore, the increase in cross-classification performance after learning was not driven by the type of association learned by participants (Figure 5): Similar increases were observed for the participants who had associated faces with cars and houses with chairs (increase of 3.4 ± 1.4%, two-tailed, paired t test; $t(9) = 2.26$, $p = .05$) as for the participants who had associated faces with chairs and houses with cars (increase of 3.2 ± 1.2%, $t(9) = 2.42$, $p < .04$).

As argued above, the shift in the distribution of cross-classification performance suggests that multivoxel object category patterns become more similar to each other after participants learn associations between the categories. How sparse are these learning-related changes? On the one hand, the representational changes could potentially be highly variable across voxels, with voxels in some areas showing a large increase in performance after learning and others showing no change at all. Alternatively, at the other extreme, every voxel in the scanned volume could shift by the same amount. To determine how specific the learning-induced changes were, we evaluated the distribution of voxelwise differences between the pre- and postlearning classification performances (Figure 4B). As expected from the results in Figure 4A, the average voxelwise increase in cross-classification performance after learning was 3.3%. However, the shift was variable: Some voxels showed an increase in cross-classification performance of more than 10%. Figure 6 shows the scatter of pre- and postlearning cross-classification performances across all voxels, and its relationship to the histograms shown in Figure 4A and B.

We next asked how the voxels that showed the largest shifts in performance were organized in cortex. In other words, did they occur together in localized regions or were they dispersed all across cortex? Some authors have suggested that expertise-related changes might occur in specific ROIs, for example, in the fusiform gyrus (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). In a first step, we thus evaluated how learning affected object representations in functionally defined regions of ventral temporal cortex that are known to be important for processing visual categories. In particular, we identified four functionally defined regions in each participant's native space: the FFA (Kanwisher et al., 1997), the PPA (Epstein & Kanwisher, 1998), and the pF and LO subdivisions of the LOC (Grill-Spector, Kourtzi, & Kanwisher, 2001). In addition, we anatomically identified the hippocampus (because of its implication in the acquisition of new associations), and the early visual cortex (V1 and V2) as a control region. In each of these areas, we performed the same analysis as in Figure 4A (Figure 4D–G). However, because we were considering smaller ROIs, we restricted this analysis to a smaller searchlight of radius 3 voxels. Note that we retained

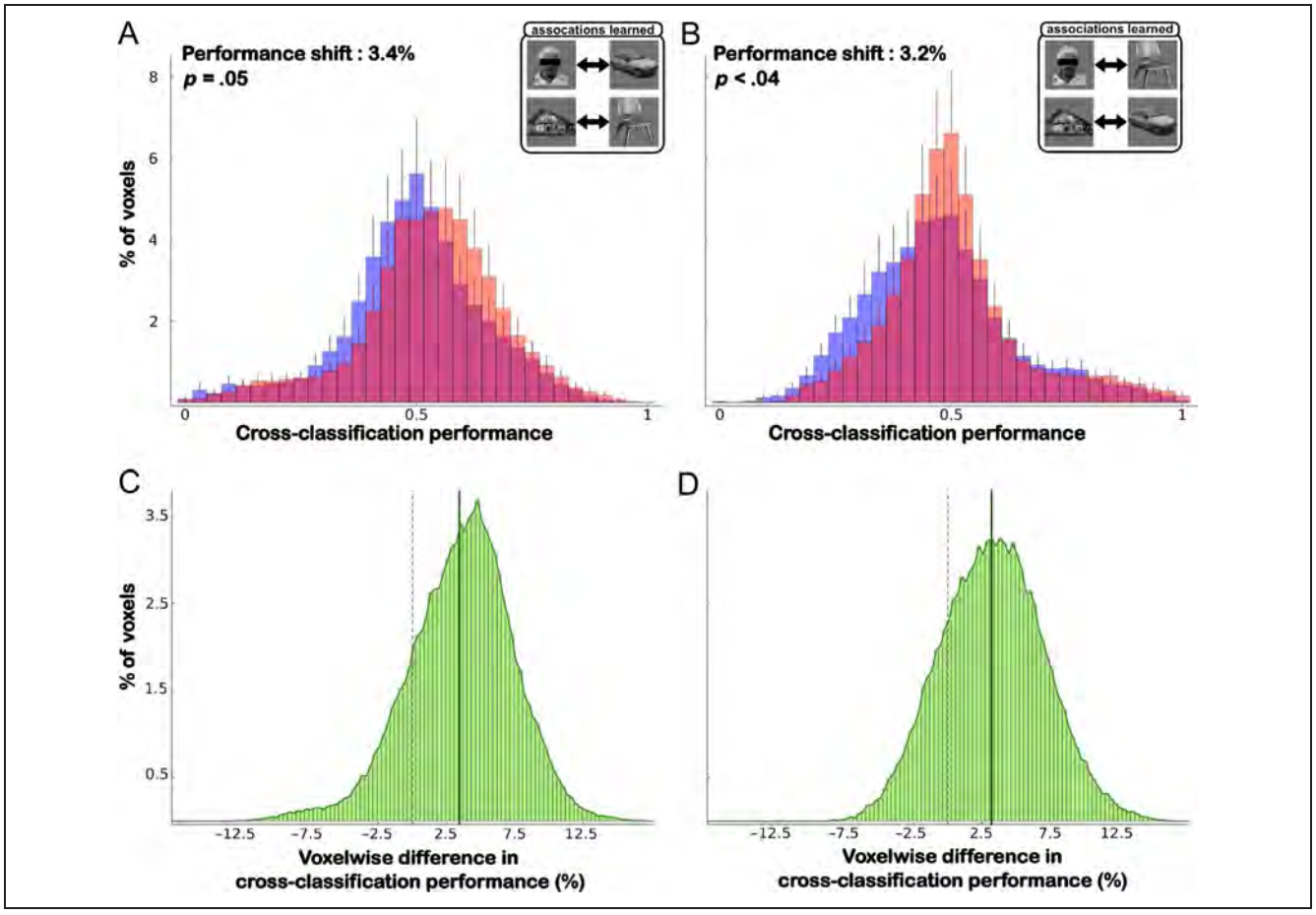

Figure 5. Cross-classification performance by association type. In our group of 20 participants, half the participants associated faces with cars and houses with chairs (Group 1; A), whereas the other half associated faces with chairs and houses with cars (Group 2; B). Both groups of participants showed similar effects of associative learning (independent samples t test, $t(9) = 0.11$, $p = .9$). A and B have the same format as in Figure 4A. C and D have the same format as in Figure 4B.

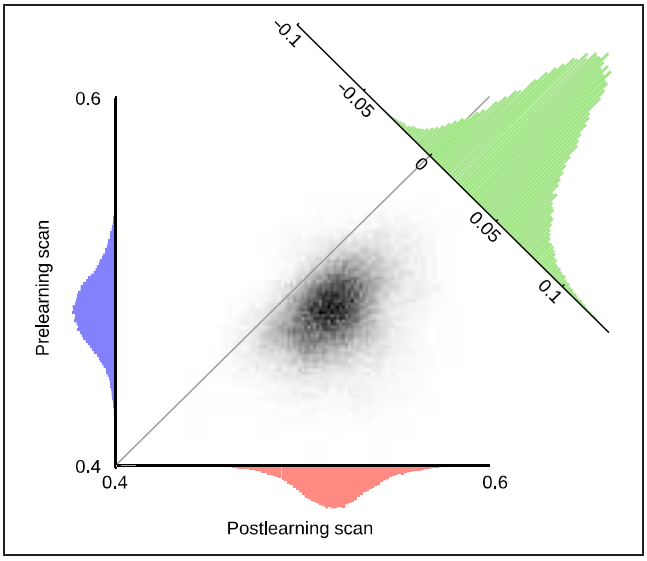

Figure 6. Scatter plot of voxelwise cross-classification performance in the prelearning versus postlearning scan sessions. The blue and pink histograms are the projections of the data on the *y*- and *x*-axes, respectively, and are similar to the data shown in the blue and pink histograms of Figure 4A (save for the fact that here the data points and corresponding histograms represent mean classification performance of each voxel across participants). The green histogram corresponds to the data in Figure 4B and is the projection of the data perpendicular to the diagonal.

a higher-resolution searchlight approach rather than testing a whole ROI classifier, because it is conceivable that over an entire ROI the most informative voxels (i.e., those that will dominate the classifier's decision) may not be those that show the strongest learning effect (and indeed, this possibility was confirmed in a subsequent analysis; see Figures 8 and 9). In that case, a whole ROI classifier may not show any learning-induced change in cross-classification (Figure 9), although individual voxels within the corresponding ROI could have significantly altered their response pattern; the searchlight method, on the other hand, would still reveal the changes in those voxels (Figure 4). We observed a statistically significant increase in cross-classification performance in all ROIs (pF: 5.6 ± 1.2%; $t(19) = 4.33$, $p < .0005$; FFA: 4.9 ± 1.1%; $t(19) = 4.24$, $p < .0005$, PPA: 3.8 ± 1.6%; $t(19) = 2.3$, $p < .05$, LO: 2.5 ± 0.8%; $t(19) = 2.99$, $p < .01$), but not in the hippocampus (0.2 ± 0.8%; $t(19) = 0.33$, $p > .7$) and V1/V2 (0.4 ± 1.2%; $t(19) = 0.72$, $p > .36$).

In a complementary analysis, we asked where the voxels that showed the largest increase in performance were localized. Figure 7 shows the average voxelwise performance differences (obtained with a searchlight of radius 3 voxels) projected on the inflated brain. To assess the statistical significance of the cross-classification shifts, we used a nonparametric test in which we shuffled the labels of the pre- and postlearning sessions for each voxel independently to simulate the null hypothesis that there was no difference between these sessions. Voxels that shifted significantly ($p < .001$, uncorrected) corresponded to a performance shift of at least 9.25% and were clustered principally in the left and right fusiform gyri. In particular, the largest group of these voxels overlapped with the functionally defined left and right FFA and an anterior subdivision of the LOC known as pF (Grill-Spector et al., 1999).

As can be seen in Figures 4 and 7, the learning-induced changes were not uniform within each ROI. Instead, some voxels exhibited greater shifts in cross-classification performance than others. We next investigated what characterized those voxels that showed higher flexibility. We reasoned that flexibility might be inversely related to initial selectivity, that is, that the voxels that originally provided the most information about object category (i.e., the most specialized voxels) would retain their selectivity, whereas the least informative voxels would be most sensitive to category associations during the learning phase. Thus, in each of the previously identified ROIs (pf, FFA, PPA, and LO), we compared the voxelwise increase in cross-classification performance after learning with the ability of that voxel to provide category-specific information before learning (i.e., the performance of a classifier trained on a set of patterns and tested on patterns from the same category). For each voxel, the performance of a four-way classifier (3-voxel radius spherical searchlight centered on that voxel), trained and tested on discriminating the four categories (faces, houses, chairs,

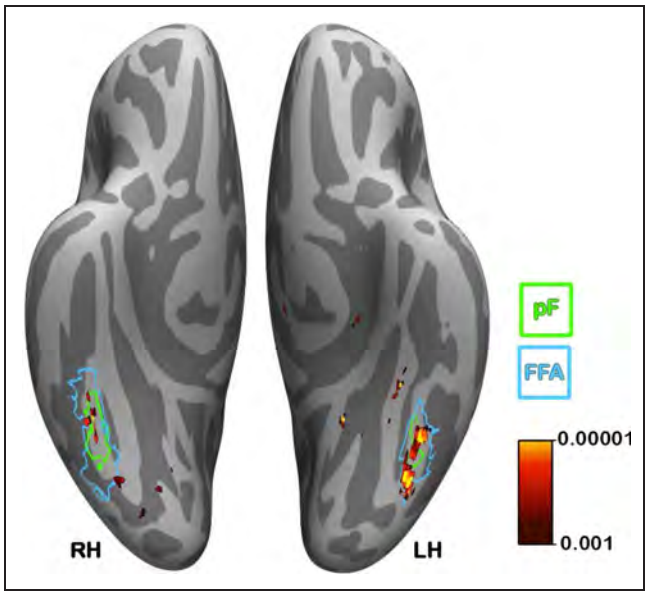

Figure 7. Localization of voxels that showed the largest increase in cross-classification accuracy. The voxels that showed the largest increase in cross-classification accuracy across all participants after learning were in relatively localized regions of the left and right fusiform gyri ($p < .001$ uncorrected; corresponding to an increase in cross-classification accuracy of 9.25% or more). The colorbar corresponds to p values (uncorrected) determined from a nonparametric test. The outline of the functionally defined FFA (averaged across participants) is shown in blue, and the average pF is shown in green.

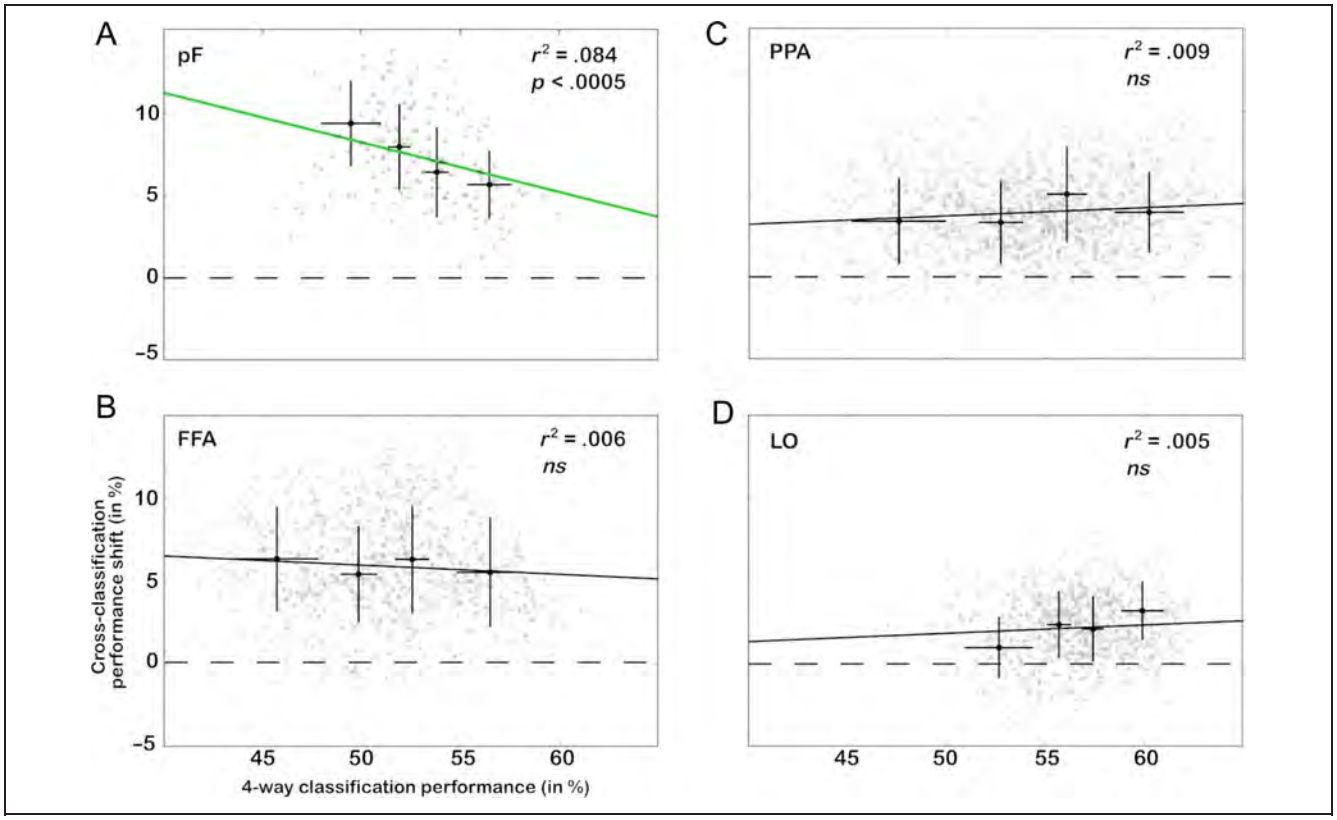

Figure 8. Voxelwise correlation of changes in cross-classification performance (Session 2–Session 1) with the initial performance of a four-way classifier. A four-way classifier (chance at 25%) was trained and tested on prelearning data to discriminate the four object categories from each other (i.e., trained on a set of patterns and tested on patterns of the same categories). Its performance was correlated with the learning-induced changes in cross-classification performance for each voxel in the (A) pF, (B) FFA, (C) PPA, and (D) LO. Increase in cross-classification performance as a result of learning was significantly negatively correlated with the category discrimination performance in area pF (p < .0005). The gray points correspond to the individual voxels in each ROI on which correlations were computed. For visibility only, the voxels were split into quartiles according to four-way classification performance. The mean performance for each quartile is shown by the black points (error bars correspond to *SD* across voxels).

cars) during the first fMRI recording session (prelearning), was correlated with the learning-induced change in cross-classification performance. Consistent with our hypothesis, increase in cross-classification performance was significantly negatively correlated with the initial performance of the four-way classifier in area pF (Figure 8). This finding indicates that, in this ROI, the voxels that exhibited the most flexibility during the learning procedure were the ones with the lowest category-specific information prelearning (albeit four-way classification performance in these voxels was much higher than the 25% chance level; Figure 8A). Note also that, although in pF

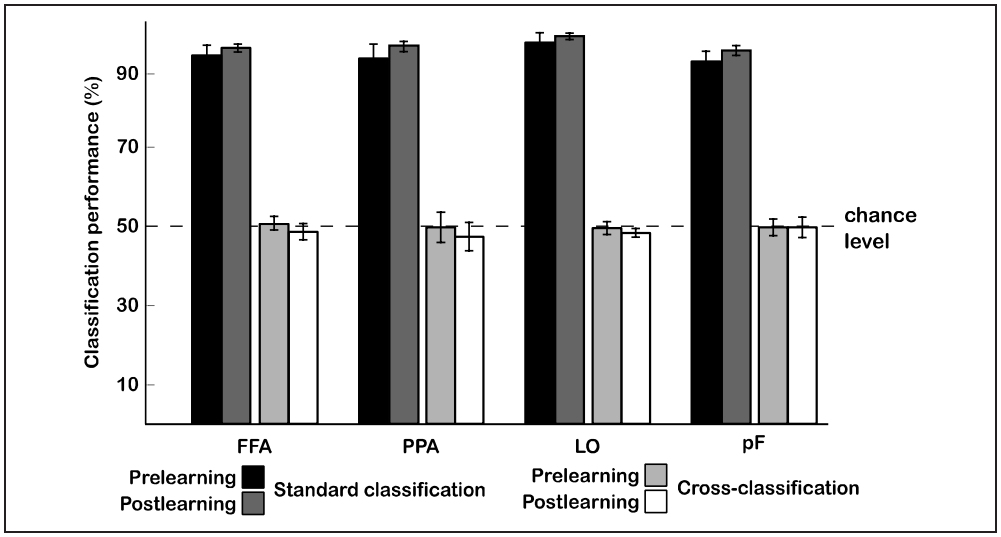

Figure 9. Prelearning and postlearning SVM classification performance in the FFA, PPA, LO, and pF, performed at the level of the entire ROI, that is, without a searchlight method. "Standard classification" refers to the average performance of a face–house (FH) classifier tested on FH discrimination and a car–chair (CC) classifier tested on CC discrimination (using a leave-one-run-out approach). "Cross-classification" refers to the average performance of the FH classifier on CC discrimination and the CC classifier on FH discrimination.

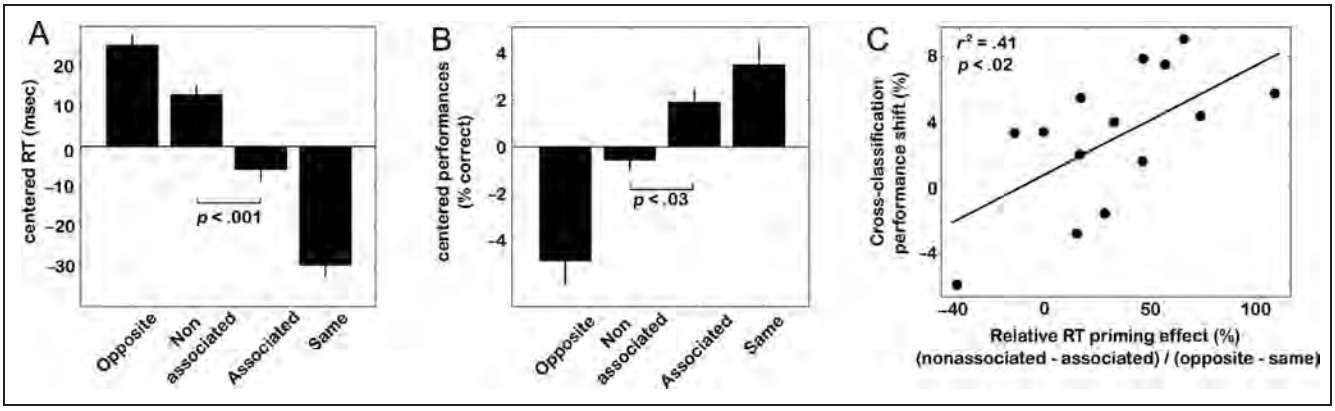

Figure 10. Behavioral performance on a priming task. Participants performed a priming task and were instructed to prioritize response times over accuracy. (A) RTs were significantly shorter (21.4 ± 5.4 msec, mean ± *SEM*, two-tailed, paired *t* test; $t(13) = 4.47$, $p < .001$) for the associated categories versus the nonassociated categories. The magnitude of the priming effect on RTs was approximately 32% of the maximal priming that could be observed between "same" and "opposite" trials (66.3 ± 5.0 msec). For each participant, the mean RT across all conditions was subtracted from each condition to obtain the centered RT displayed here. For statistical tests only (but not for display), the RT data were log-transformed to satisfy the constraints of normality. (B) Although RTs were our main dependent variable, a compatible difference was also observed for accuracies on "associated" versus "nonassociated" trials (2.5 ± 0.9%, two-tailed, paired *t* test; $t(13) = 2.53$, $p < .03$). (C) The priming effect on RTs (expressed relative to the maximal priming) was statistically correlated over the group of participants with the difference in cross-classification performance before and after learning ($r^2 = .41$; $p < .05$).

the voxels with the least category selectivity showed the largest learning effects, at the level of the entire ROI pF itself was highly category selective (Figure 9).

One may question the validity of using a local selectivity measure (the searchlight method) to draw global conclusions about the entire ensemble of recorded voxels across occipital and temporal cortex: global measurements (such as a classifier trained and tested on the entire set of voxels) may appear more appropriate for that purpose. In fact, however, the searchlight method allowed us to obtain a global measure of learning over the whole brain (by averaging across voxels) and to then subsequently hone in on more localized effects. Note that, as opposed to this approach of the searchlight method, a classification analysis performed over all the voxels in the entire scanned volume (or even in a specific ROI, as alluded to above; see Figure 9) could potentially fail to find the voxels that show the biggest changes in cross-classification. For example, a global classifier trained to discriminate faces versus houses across a large swath of cortex would identify the voxels that are the most informative (i.e., category selective) for this face/house discrimination task and disregard the voxels that are the least category selective. However, as we observed in Figure 8, the voxels that were the most prone to learning in pF (i.e., showing the most significant learning effects in a cross-classification task) were precisely the ones that were the least category selective. Thus, although a whole-brain classifier might assign negligible weights to these voxels and consequently fail to identify learning effects, the searchlight method would not because it is constrained to learn from local patterns.

Do the multivoxel representational shifts have perceptual consequences at the behavioral level? In a priming task, performed ~14 months after the associative learning had occurred, we investigated whether perception of one category facilitated the behavioral processing of its associated category, relative to its nonassociated category (Figure 2). An examination of participants' behavioral performance revealed that RTs were significantly shorter (two-tailed, paired *t* test; $t(13) = 4.47$, $p < .001$) on trials when the prime stimulus was a paired associate versus a nonassociate (Figure 10A). The average magnitude of this priming effect (21.4 ± 5.4 msec, mean ± *SEM*) was approximately 32% of the maximal priming (66.3 ± 5.0 msec, mean ± *SEM*) that could be observed between "same" and "opposite" trials. Although RTs were our main dependent variable (because participants were explicitly instructed to prioritize response speed over accuracy), a compatible difference was also present for accuracies on "associated" versus "nonassociated" trials (2.5 ± 0.9%, two-tailed, paired *t* test; $t(13) = 2.53$, $p < .03$), with a priming effect for associated categories that was 22.5% of the corresponding maximal priming (Figure 10B).

Could this priming effect represent a behavioral correlate of the cortical representational shifts observed in the fMRI? In support of this idea, we found that the priming effect on RTs (expressed relative to the maximal priming) was statistically correlated over the group of participants with the difference in cross-classification performance before and after learning ($r^2 = .41$ $p < .05$; 95% confidence interval: $.07 \leq r^2 \leq .75$; Figure 10C). In other words, the participants who had displayed the maximal shifts in multivoxel representations were also those who showed the largest priming effects. Thus, we found that even several months after the associative learning had occurred, changes in neural representations of the associated categories were accompanied by significant and commensurate response priming at the behavioral

level (although, as with all effects based on a correlation analysis, these results cannot provide evidence for a direct link between the changes observed in fMRI and the behavioral priming effects).

DISCUSSION

In this study, we asked how associative learning changes large-scale multivoxel representations in ventral temporal cortex. After learning, we observed an average increase of 3.3% in cross-classification performance of multivoxel category representations, with some voxels showing shifts of up to 10%. Because our experiment used a block design, it remains an open question whether these multivoxel category patterns arise spontaneously in the brain under different testing regimes (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008). Nevertheless, our results suggest that in conditions where category-specific MVPs can be recorded, the multivoxel representations for associated categories in object-selective cortex become more similar to each other after associations are learned. In a behavioral experiment, we verified the perceptual consequences of the shifts in multivoxel representations several months after the learning had occurred. Not only did paired associates produce significant cross-category priming, but also, the participants who had displayed the maximal shifts in multivoxel representations were those who showed the largest priming effects. Note however that we cannot exclude other factors that might also have contributed to the significant correlation between fMRI effects and behavior, for example, participants' motivation levels and their ability to follow task instructions.

Cross-classification performance after learning was significantly higher than before learning. However, when averaged over all participants and voxels (Figure 4), neither the pre- nor postlearning cross-classification performance values (48.3% and 51.7%, respectively) were significantly different from chance level (50%; $t(19) = 1.55$, $p = .13$, for the prelearning distribution and $t(19) = 1.82$, $p = .08$, for the postlearning distribution). We believe that the initially low performance value was caused by spontaneous biases in category associations occurring in many brain areas. In the PPA, for example, on average about 62% of the chair-category patterns tended to be spontaneously associated with house (rather than face) patterns and cars with faces (rather than houses). Of course, the counterbalanced set of participants was designed to minimize the effects of any such initial bias (because for one half of the participants, this bias would result in lower-than-chance prelearning cross-classification and higher than chance for the other half). However, in our limited participant population, it is not altogether surprising that the initial bias of a few participants could have been overly represented in the grand average (e.g., because of a higher signal-to-noise ratio during scanning or because the relative volume of specific ROIs was bigger in these participants), leading (in our case) to an average prelearning cross-classification below 50%. If we take this initially low value as the chance level (or baseline) for postlearning cross-classification, therefore, we observe a truly significant ($p < .003$) cross-classification improvement due to learning. It must also be emphasized that our findings are not contingent on below-chance prelearning cross-classification: Similar learning-induced improvements were registered for several brain regions and participants whose cross-classification accuracy started off above chance. This can be easily visualized in Figure 6: Even the voxels with the highest initial cross-classification performance demonstrated a learning-induced improvement (i.e., a shift to the right of the diagonal).

The shifts observed in the MVPs could reflect different mechanisms by which object representations change as a result of learning. For instance, the new patterns could reflect a link (or a coactivation) between the two (unchanged) initial representations of the associated categories or signal entirely new representations that combine information about the associated categories. Although it would be interesting to compare what category information is encoded in the initial versus changed representations, we must note that any comparison of MVPs across the two sessions (i.e., training on patterns from one session and testing on patterns from the other) would confound learning effects with pattern and classification differences that are simply due to the fact that the two scan sessions were obtained on different days. However, the finding that, in area pF, the voxels that showed the greatest flexibility during learning were the ones that originally provided the least (albeit still much greater than the 25% chance level) category-specific information (Figure 8) suggests that the voxels that are the most informative in encoding category information mostly preserve their response profiles whereas the least informative voxels are more readily modulated by associative learning.

Object representations in the VVP can be described both at the level of individual neuronal selectivities as well as in large-scale multivoxel activation patterns (Reddy & Kanwisher, 2006). Indeed, in the human brain, MVPs are often used as a proxy for understanding the neuronal codes underlying object representation (Stansbury, Naselaris, & Gallant, 2013; Kriegeskorte, Mur, Ruff, et al., 2008; Haynes & Rees, 2005; Kamitani & Tong, 2005; Carlson, Schrater, & He, 2003; Spiridon & Kanwisher, 2002; Haxby et al., 2001). As explained earlier, the observed increase in cross-classification performance after learning can be described in mathematical terms as a shift of the MVPs in a high-dimensional multivoxel space. However, this shift of MVPs could arise from different mechanisms at the neuronal level, and we can only speculate here about such neuronal properties. For instance, individual neurons within each voxel could change their tuning curve as a result of learning, such that initially face-selective neurons (for example) would also now respond to the associated chairs (Ison et al., 2015; Reddy et al., 2015). Such a

change in tuning is equivalent at the neuronal level to the coactivation account alluded to above. That is, when participants view one stimulus (e.g., a chair), neurons that are normally selective to the associated stimulus (e.g., a face) could also be partially and automatically activated, occasioning a change of their tuning curve. In turn, this would imply that the recorded MVP in response to viewing a chair is composed of a combination of chair and face representations. Alternatively, the newly learned associations could be encoded within each voxel by a new set of neurons that were previously nonselective for either stimulus of the associated pair. In other words, when viewing a face or a chair, in addition to the original populations of face- and chair-selective neurons (respectively), a new subpopulation of neurons encoding the face–chair relation would also be activated. Although our data do not allow us to tease apart these different (and nonexclusive) mechanisms at the neuronal level, they do provide evidence that object representations as measured by MVPs are not static. Recent studies have shown that multivoxel representations of objects in ventral temporal cortex are not fixed but can be modulated by top–down signals such as task goals (Harel, Kravitz, & Baker, 2014). Our findings add to this body of work and show that object representations of highly familiar categories can flexibly move in a high-dimensional multivoxel space as a result of associative learning.

During tasks of explicit memory recall, when participants learn to pair two stimuli together (e.g., a word and a scene), the presentation of a cue stimulus (e.g., the word) can reactivate the fMRI representation of the associated stimulus (Gordon, Rissman, Kiani, & Wagner, 2014; Kuhl & Chun, 2014; Kuhl, Rissman, Chun, & Wagner, 2011; Johnson, McDuff, Rugg, & Norman, 2009; Polyn, Natu, Cohen, & Norman, 2005; Nyberg, Habib, McIntosh, & Tulving, 2000; Wheeler, Petersen, & Buckner, 2000). This reactivation of associated stimuli during explicit recall appears to resemble the results reported here and is compatible with the coactivation account discussed above. Note, however, that this study differs in one crucial aspect from past studies of explicit recall. In the studies cited above, the reactivation of the associated MVP occurred as the participants were explicitly instructed to perform a recall task (and thus retrieve the corresponding stimulus in memory). In contrast, in our study, participants were not instructed to perform a recall task of associated stimuli. Instead they performed a 1-back task on the currently viewed images that was independent of any recall or associative learning. The changes in fMRI representations were observed while participants performed this 1-back task and in the presence of competing visual stimuli (e.g., information about chair stimuli could be decoded while participants were actually viewing and performing a task on faces). Thus, although we cannot discount the possibility that participants automatically recalled a chair while viewing the associated face, this recall must necessarily have occurred in the presence of competing visual input and simultaneously with the performance of a nontrivial, independent task performed on the currently perceived stimuli and that did not require explicit recall. In the end, as discussed above, although such automatic recall could be one of the possible mechanisms underlying the increase in cross-classification performance in our experiment, it is still consistent with the conclusion that multivoxel object representations can be flexibly modified through associative learning.

The finding that the largest learning-dependent changes (>9% increase in cross-classification performance) were observed in clusters of voxels in the left and right fusiform gyri is consistent with a previous study showing associative learning effects in the left fusiform cortex (Park, Shannon, Biggan, & Spann, 2012). The voxels showing the largest changes overlapped substantially with our functionally defined FFA, as well as with an anterior subdivision of the LOC located in the fusiform gyrus (pF; Grill-Spector et al., 1999). The object-selective pF itself partially overlapped with the FFA (Grill-Spector et al., 2001), but we were unable to further segregate these two ROIs in the native space of each participant. Other recent studies have also reported a mix of face- and object-selective voxels in the traditionally defined FFA (Cukur, Huth, Nishimoto, & Gallant, 2013; Hanson & Schmidt, 2011). It has been argued that increased expertise with a class of objects is correlated with the level of activation in the FFA (McGugin, Gatenby, Gore, & Gauthier, 2012; Gauthier et al., 1999), although this claim is still debated (McKone, Kanwisher, & Duchaine, 2007; Kanwisher, 2000). Although our data are unable to shed light on this debate because of the spatial overlap between the FFA and the pF, we find that face- and object-selective representations in the fusiform gyrus show the strongest changes in representational similarity as a result of associative learning.

Previous studies have investigated the effects of training on object representations in object-selective cortex. In general, these studies reveal that training-related changes occur in a distributed fashion in inferotemporal cortex and that these changes are often modest (Op de Beeck & Baker, 2010). In monkeys, training changes the selectivity and strength of neuronal responses in inferotemporal cortex (Li & DiCarlo, 2008; Freedman, Riesenhuber, Poggio, & Miller, 2006; Baker, Behrmann, & Olson, 2002; Sigala & Logothetis, 2002; Logothetis, Pauls, & Poggio, 1995). Human fMRI studies have shown that learning is associated with increases or decreases in the overall amplitude of the average BOLD response (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006; Kourtzi, Betts, Sarkheil, & Welchman, 2005; Gauthier et al., 1999), as well as with a sharpening of neural tuning (Zhang, Meeson, Welchman, & Kourtzi, 2010; Gillebert, Op de Beeck, Panis, & Wagemans, 2009; Jiang et al., 2007). The current study extends this previous work by investigating the effects of associative learning on preexisting, well-established response patterns for pairs of familiar categories (rather than extensive practice with a single category).

Previous monkey studies have investigated a class of neurons known as "pair-coding neurons" that respond similarly to pairs of stimuli that have been associated together (Sakai & Miyashita, 1991). In these studies, monkeys learned associations between novel, meaningless fractal patterns that they had been exposed to on a relatively short timescale (i.e., in recent experimental sessions). After learning, a neuron that was originally selective to a cue stimulus showed selective responses to its paired associate as well. However, neuronal selectivity for novel stimuli (e.g., the cue stimuli in the aforementioned studies) can flexibly develop as a result of recent exposure (Logothetis et al., 1995), suggesting that the pair-coding task principally modified neuronal responses in recently created representations. In contrast, our participants learned novel associations between already overlearned categories of stimuli, with which they had lifelong exposure. After learning, we found that category selectivity was modified in well-established (and hence presumably less flexible) multivoxel representations that are thought to contribute to visual categorization and object representation. Additionally, pair-coding neurons show significantly correlated responses to pairs of pictures (i.e., at the exemplar level) in a stimulus–stimulus association task. In contrast, we found that category level multivoxel representations change, although the associations were created between exemplars of the two categories. Finally, pair-coding neurons have typically been found in the anterior ventral portion of area TE and in the perirhinal cortex (although a larger proportion of these neurons and stronger pair-coding effects were found in the perirhinal cortex; Naya, Yoshida, & Miyashita, 2003). Other studies have also found evidence for associative learning in perirhinal cortex and anterior ventral IT neurons in monkeys (Eifuku, Nakata, Sugimori, Ono, & Tamura, 2010; Erickson & Desimone, 1999) and in single neurons in the human MTL (Ison et al., 2015; Reddy et al., 2015). However, information about associated stimuli has not been found in single neurons in more posterior portions of TE (Gochin, Colombo, Dorfman, Gerstein, & Gross, 1994). In this study, we observed the strongest effects of associative learning in voxels in the fusiform cortex, overlapping with the FFA and pF. Although it is difficult to establish exact homologies between the monkey and human brains, the human LOC and FFA are thought to correspond to the posterior and dorsal part of the monkey inferotemporal complex (Tsao, Moeller, & Freiwald, 2008; Denys et al., 2004). Our findings thus suggest that information about associated stimulus pairs is also observed in human visual regions more caudal to those previously reported in single neurons in monkey anterior ventral inferotemporal cortex.

Acquiring new associations depends critically on MTL structures, including the hippocampus (Squire, Stark, & Clark, 2004; Fortin, Agster, & Eichenbaum, 2002). As mentioned above, single-neuron recordings in monkeys (Wirth et al., 2003; Erickson & Desimone, 1999; Sakai & Miyashita, 1991; Miyashita, 1988) and humans (Ison et al., 2015; Reddy et al., 2015) show that MTL neurons change their selectivity as a result of learning associations between pairs of stimuli. Human fMRI studies have implicated different MTL structures in associative learning, sequence learning, and relational memory (Schapiro, Kustner, & Turk-Browne, 2012; Turk-Browne, Scholl, Chun, & Johnson, 2009; Haskins, Yonelinas, Quamme, & Ranganath, 2008; Aminoff, Gronau, & Bar, 2007; Diana, Yonelinas, & Ranganath, 2007; Davachi, 2006; Prince, Daselaar, & Cabeza, 2005). In particular, hippocampal fMRI activity patterns become more similar to each other as a result of incidental sequence learning (Schapiro et al., 2012). In our study however, object category multivoxel representations in the hippocampus were essentially unmodified during the postlearning scan. This difference between the two studies could be accounted for by differences in the learning protocols. For instance, in the previous study participants viewed sequences of items but were unaware of the relationships between them. In our study however, participants were explicitly instructed to make associations between the object categories. Additionally, in our study learning occurred over a much longer time frame, with the result that the associations were overlearned (Figure 3) when postlearning brain activity was measured. Thus, although the hippocampus undoubtedly plays an active role during the acquisition of new associations, for instance by differentially activating for successfully learned versus unlearned associations (Davachi, 2006), it is possible that the relevant information was processed and stored in other cortical areas once the associations were overlearned. Indeed, although it is not known how long memory traces need to remain active in MTL structures before being committed to long-term storage in anterior inferotemporal cortex, the representational changes we observe in the VVP could be consistent with such a reorganization of learned information.

Participants were explicitly asked to learn arbitrary associations between unrelated object categories, and we measured changes in neural response patterns in an fMRI scan session at the end of learning. Learned associations in this case could be direct and automatic or mediated by explicit strategies such as recall (as described earlier) and/or visual imagery. Visual perception and visual imagery of familiar categories of objects have been shown to elicit similar patterns of fMRI activity in ventral visual cortex (Reddy, Tsuchiya, & Serre, 2010). Recall of past visual stimuli also reactivates their representations in visual cortex (Wheeler et al., 2000). It is conceivable that during the postlearning scan of the current study, while viewing one category of images (e.g., chairs), participants brought the associated category (e.g., faces) to mind, although they performed a 1-back task on the images that was independent of any associative learning. However, note that even if participants could not avoid recall and mental imagery of the associated categories, the very experience of a stimulus "bringing another to mind" when the task (1-back)

does not require such recall is arguably a manifestation of a well-learned association.

To conclude, we show that associative learning is accompanied by large-scale neural changes in the VVP. Specifically, multivoxel activity patterns for associated object categories become more similar to each other with learning. An interesting open question that we have not addressed here is whether these representational changes are specific to the stimuli with which learning occurred, or whether they generalize to other exemplars in the category. Additionally, how long do these changes persist after the learned associations are no longer behaviorally relevant? Although these questions remain exciting topics for future research, here we show evidence for flexible and dynamic representations in ventral temporal cortex that could support the daily process of learning new relationships between different events.

Acknowledgments

This work was supported by funding from an ANR-JCJC (2012) to L. R. and funding from the Institute des Sciences du Cerveau de Toulouse to L. R. and R. V. We thank Francisco Pereira for sharing his SearchMight toolbox with us. L. R. and R. V. designed the research. The authors would like to thank the staff of the Imaging Center, INSERM/UPS UMR 825 MRI platform for their assistance in acquiring the data.

Reprint requests should be sent to Dr. Leila Reddy, CNRS-Centre de Recherche Cerveau et Cognition, Pavillon Baudot CHU Purpan, 31052 Toulouse Cedex, France, or via e-mail: leila.reddy@cerco.ups-tlse.fr.

REFERENCES

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex, 17,* 1493–1503.

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience, 5,* 1210–1216.

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience, 15,* 704–717.

Cukur, T., Huth, A. G., Nishimoto, S., & Gallant, J. L. (2013). Functional subdomains within human FFA. *Journal of Neuroscience, 33,* 16748–16766.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology, 16,* 693–700.

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., et al. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: A functional magnetic resonance imaging study. *Journal of Neuroscience, 24,* 2551–2565.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences, 11,* 379–386.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473.

Eifuku, S., Nakata, R., Sugimori, M., Ono, T., & Tamura, R. (2010). Neural correlates of associative face memory in the anterior inferior temporal cortex of monkeys. *Journal of Neuroscience, 30,* 15085–15096.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Erickson, C. A., & Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. *Journal of Neuroscience, 19,* 10404–10416.

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience, 5,* 458–462.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex, 16,* 1631–1644.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience, 2,* 568–573.

Gillebert, C. R., Op de Beeck, H. P., Panis, S., & Wagemans, J. (2009). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. *Journal of Cognitive Neuroscience, 21,* 1054–1064.

Gochin, P. M., Colombo, M., Dorfman, G. A., Gerstein, G. L., & Gross, C. G. (1994). Neural ensemble coding in inferior temporal cortex. *Journal of Neurophysiology, 71,* 2325–2337.

Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cerebral Cortex, 24,* 3350–3364.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13,* 159–166.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41,* 1409–1422.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24,* 187–203.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27,* 649–677.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics, 7,* 37–53.

Hanson, S. J., & Schmidt, A. (2011). High-resolution imaging of the fusiform face area (FFA) using multivariate non-linear classifiers shows diagnosticity for non-face categories. *Neuroimage, 54,* 1715–1734.

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences, U.S.A., 111,* E962–E971.

Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron, 59,* 554–560.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., et al. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron, 72,* 404–416.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8,* 686–691.
Ison, M. J., Quian Quiroga, R., & Fried, I. (2015). Rapid encoding of new memories by individual neurons in the human brain. *Neuron, 87,* 220–230.
Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron, 53,* 891–903.
Johnson, J. D., McDuff, S. G., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron, 63,* 697–708.
Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8,* 679–685.
Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3,* 759–763.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.
Kourtzi, Z., Betts, L. R., Sarkheil, P., & Welchman, A. E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. *PLoS Biology, 3,* e204.
Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.
Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 4.
Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141.
Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12,* 535–540.
Kuhl, B. A., & Chun, M. M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *Journal of Neuroscience, 34,* 8051–8060.
Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences, U.S.A., 108,* 5903–5908.
Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science, 321,* 1502–1507.
Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5,* 552–563.
McGugin, R. W., Gatenby, J. C., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proceedings of the National Academy of Sciences, U.S.A., 109,* 17063–17068.
McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11,* 8–15.
Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 12239–12244.
Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature, 335,* 817–820.
Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331,* 68–70.
Naya, Y., Yoshida, M., & Miyashita, Y. (2003). Forward processing of long-term associative memory in monkey inferotemporal cortex. *Journal of Neuroscience, 23,* 2861–2871.
Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences, U.S.A., 97,* 11120–11124.
Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in Cognitive Sciences, 14,* 22–30.
Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience, 26,* 13025–13036.
Op de Beeck, H. P., Brants, M., Baeck, A., & Wagemans, J. (2010). Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. *Neuroimage, 49,* 3414–3425.
Park, H., Shannon, V., Biggan, J., & Spann, C. (2012). Neural activity supporting the formation of associative memory versus source memory. *Brain Research, 1471,* 81–92.
Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310,* 1963–1966.
Prince, S. E., Daselaar, S. M., & Cabeza, R. (2005). Neural correlates of relational memory: Successful encoding and retrieval of semantic and perceptual associations. *Journal of Neuroscience, 25,* 1203–1210.
Reddy, L., & Kanwisher, N. (2006). Coding of visual objects in the ventral stream. *Current Opinion in Neurobiology, 16,* 408–414.
Reddy, L., & Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Current Biology, 17,* 2067–2072.
Reddy, L., Poncet, M., Self, M. W., Peters, J. C., Douw, L., van Dellen, E., et al. (2015). Learning of anticipatory responses in single neurons of the human medial temporal lobe. *Nature Communications, 6,* 8556.
Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *Neuroimage, 50,* 818–825.
Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature, 354,* 152–155.
Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology, 22,* 1622–1627.
Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature, 415,* 318–320.
Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35,* 1157–1165.
Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience, 27,* 279–306.
Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron, 79,* 1025–1034.
Straw, A. D. (2008). Vision egg: An open-source library for realtime visual stimulus generation. *Frontiers in Neuroinformatics, 2,* 4.

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 19514–19519.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21,* 1934–1945.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, U.S.A., 97,* 11125–11129.

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science, 300,* 1578–1581.

Zhang, J., Meeson, A., Welchman, A. E., & Kourtzi, Z. (2010). Learning alters the tuning of functional magnetic resonance imaging patterns for visual forms. *Journal of Neuroscience, 30,* 14127–14133.

C. Discussion

1. Summary

In this study we investigated the flexibility of visual object representations in the brain by comparing their multi-voxel activation patterns before and after an associative learning protocol of three weeks. Our results show that neural representations of associated categories were modulated by the learning protocol, more specifically the representations of associated categories became more similar as a result of learning. This shift was the strongest in the left fusiform gyrus, more specifically in the posterior fusiform. Moreover, when examining the shift in multi-voxel representations of associated visual categories in left and right posterior fusiform gyri, voxels that had a strong selectivity before learning showed the smallest learning induced changes, and vice versa. In other words the associative learning effect was negatively correlated with the pre-learning selectivity to the four categories, which suggests that neuronal populations which did not show a strong category specific responses were more subject to incorporate the learned associations. Finally, using a visual priming experiment, we were able to show that the overall amount of shift in visual category multi-voxel representations was correlated with facilitation in reaction times and accuracy in a categorization task if subjects were primed with the associated category in contrast to the non-associated category.

To summarize, this study shows that even existing neural representations of visual object categories are flexible and can be modulated through experience, that the strongest modulation occur in the left fusiform gyrus and that the amount of modulation of object representations is correlated with behavior.

2. The left fusiform gyrus

In our study we found that the shift in representations was the largest and most significant in the left fusiform gyrus, more specifically in the posterior part of the fusiform gyrus (pFus). Studies investigating the neural basis of reading have consistently found a portion of the neighboring occipito-temporal sulcus (OTS), see Figure 25, namely the VWFA (cf. Chapter I), to be implicated in different

aspects of reading acquisition (Cohen et al., 2000). This area is particularly responsive to words compared to other visual object categories and even pseudo-words (i.e. strings not belonging to the language lexicon), strings of non-letter symbols, and activation in this region correlates with accuracy in a speeded letter recognition task (McCandliss et al., 2003). This proximity between the locus of our maximal effect and the reported VWFA might be due to a greater plasticity of the occipito-temporal area, specifically nearby the fusiform gyrus, in comparison with its right hemisphere topographical homolog, although the main hypothesis for the left lateralization of the VWFA is that language related areas are most often located ipsi-laterally (for oral comprehension and production).

Another possible reason for the prominence of our effect in the left pFus might be in line with some studies which suggested that object representations in the right hemisphere might be more specific than in the left hemisphere where these representations would be more abstract and represent categorical information rather than individuation of category exemplars, which fits with accounts of the right FFA to be an expertise area recruited when subjects become experts at individuating exemplars in certain category, such as faces for most humans or birds for ornithologists (Gauthier et al., 1999). Our analysis compared category specific multi-voxel patterns to test for category-level shifts in neural representations, and following the hypothesis of the left hemisphere being more invariant to exemplar level differences, such as letter fonts or size for visual words (Vuilleumier et al., 2002).

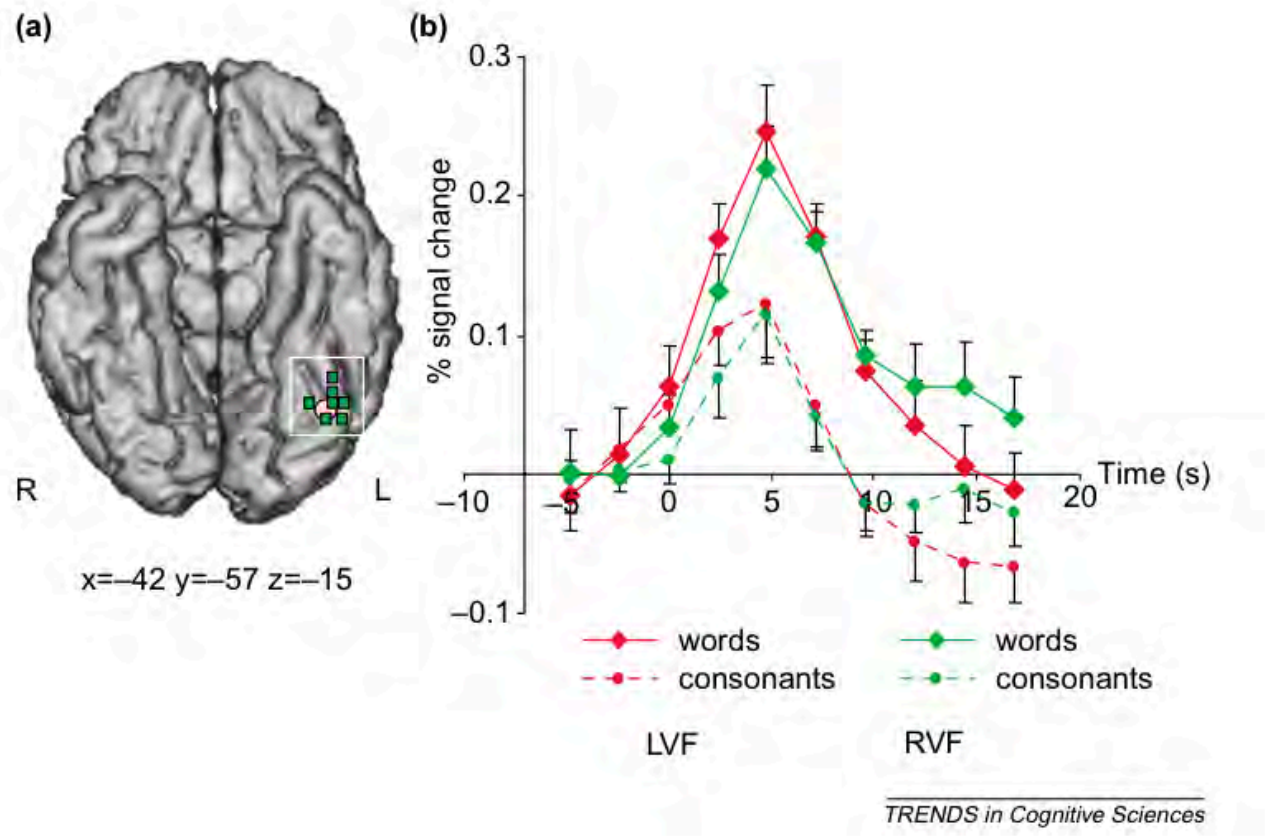

Figure 25 Peak activation point across subjects for different conditions for stimulus presentation in left or right visual field, respectively LVF and RVF (from (McCandliss et al., 2003))

3. Perspectives

In this study we showed that learning association between stimuli by category pairs (e.g. faces and cars) the multi-voxel patterns they evoked shifted and became more similar. As a follow-up experiment to test the generalization effect to other exemplars of the associated categories it would be interesting to:

1. Have different sets of stimuli for training and fMRI sessions
2. Or, to push the idea further, train subjects using different exemplars of each category at every session while keeping the category pairing the same (e.g. faces<->cars and chairs<->houses) and use either a subset of them at the pre and post scans or again new exemplars from the same categories
    a. That way we would force the system to learn categorical associations repeatedly and avoid any exemplar specific effect

b. And it would permit us to mix exemplars from the training and new ones in the scanner to see what is the extent of generalization, measured by how much new exemplars shift towards the associated category relative to exemplars which have been associated before

We used a block design protocol to increase the signal-to-noise ratio of the category-specific multi-voxel patterns. Using an event-related paradigm in a follow up experiment would allow us to ask what is exemplar-specific and what is category-specific in the multi-voxel patterns of each exemplar.

The priming experiment was done at the exemplar-association level but further it would also be interesting to investigate if the perceptual effects generalized. For example if a subject learned face1<->car3, our priming experiment tested this specific priming pair but by testing face1 as a prime and car2 (or any other than car3) it would be possible to test how the perceptual consequences of the long-term associative learning generalizes to other exemplars.

Chapter IV: Discussion, conclusion and perspectives

A. Summary

1. Aim of the thesis

The aim of this thesis was to explore the flexibility of the visual system's object representations along two of its dimensions (time and space) and at short and long time periods. The way the visual system is molded by the environment is nowadays of well acknowledged importance, from orientation bias in early visual cortex leading to perceptual bias (Girshick et al., 2011) to reorganization of functional regions in the infero-temporal lobe, more specifically in the ventral visual pathway (Dehaene et al., 2015; Grill-Spector et al., 2008). One very popular way to study neural representations that has been growing in the past 15 years is Multivariate Pattern Analyses (MVPA). These techniques allow to use the multivariate nature of the neural signal to improve the amount of information that can be extracted from neuroimaging (Harrison and Tong, 2009) and represents a more ecological framework to apprehend how the brain gives rise to behavior because of the networking and interactive nature of its structures.

Using these techniques we aimed at answering a few questions: how does learning shape visual object neural representations? How are the first stages of first stages of visual processing impacted by associative learning between stimuli? Which brain structures are actors or impacted by the integration of this new information? What are the behavioral consequences of these changes?

Each of these questions can be asked in a temporal **AND** spatial way but because of time limitations (among other things) we chose to answer these questions by building two experiments, one over a short time scale focusing on the temporal dynamics of visual processing and their flexibility when learning a sequence of visual stimuli and a second experiment at a longer timescale on the long-lasting effects on the neural functional structure and behavior of over-learning associations between existing visual categories.

2. Main results

In the first experiment we used EEG recordings during a visual sequence-learning paradigm. The results of that experiment suggest that in a context where the environment is predictable and thus permits to create expectations about the next-to-come input, e.g. navigation, the brain generates representations that are selective to the expected event in its temporally fixed spatial structure (scalp distribution) and in its spatially fixed temporal structure (time course of activity at an electrode). More specifically these spontaneously generated representations are present in the absence of the expected event at a very early latency and in specific frequency bands (high-alpha/low-beta bands). In the spatial domain the informative electrodes were located in centro-frontal, right temporal and occipital sites. The centro-frontal electrodes suggest a top-down mechanism from a hippocampal and/or frontal system generating the expectation, as observed in top-down modulation of sensory cortices in relatively related cognitive processes such as working-memory (Gazzaley and D'Esposito, 2007; Gazzaley and Nobre, 2012) and spatial navigation (Mitchell et al., 2008). The more posterior electrodes suggest a reactivation of the sensory system hosting the representations of the expected input (Gazzaley and Nobre, 2012). At a longer timescale a simple Hebbian framework would predict that the links between the co-occurring representations would be strengthened, but even though long-lasting effects of visual learning have been investigated when experiencing new visual object categories (Beeck et al., 2006) it was not fully understood how these effects would interact with existing neural representations of visual objects.

To tackle this question we realized an fMRI experiment where subjects were scanned before and after a three--week associative learning protocol. The goal was to compare neural representations of associated object categories before and after a long-term associative learning paradigm. The most prominent result was that neural representations of visual object categories which had been experienced (almost) from birth like faces, cars, chairs and houses were modulated by the associative learning paradigm. They "shifted" and associated categories got more similar to each other, e.g. multi-voxel patterns of faces became more similar to patterns of cars if a subject learned face<->car

associations. When measuring the maximum peak of this distributed shift we found the left posterior fusiform cortex, a region neighboring the VWFA, a functional region that is shaped by learning to read. This finding thus poses an interesting question on the plastic properties of this region of the ventral temporal cortex. The changes in neural representations for associated categories was also correlated with the pre-learning category classification performance, in other words the less a region was selective to the 4 categories initially the more shift it showed. This result indicates that parts of the VVP which were not very selective to any category shifted the most and it could be interpreted as a sign that a new representation, of both items of a pair, e.g. faces and cars, has emerged in the VVP organization. Finally, to investigate how a reorganization of such an important part of the neural substrate of visual processing impacted perception, subjects came back to participate in a follow-up priming experiment on average 14 months later and showed facilitation effects on a categorization task when primed with the associated category. Furthermore the magnitude of the facilitation in reaction times was significantly correlated with the neural effect observed in the function MRI data. This study showed that even early neural representations of visual categories such as faces or objects like cars, which are thought to be created by years of experience (Grill-Spector et al., 2008) in contrast with fast learning effects observed at the end of the VVP near the MTL (Reddy et al., 2015), can be modulated by experience and that the modulations we observed were behaviorally relevant because of the correlation between behavioral and neural effects.

B. Can we talk about our representation of neural representations?

Whether with short-term associative learning experiment with EEG or with the long-term MRI experiment: **what** is the difference between changes in a specific / dedicated representation, i.e. the representation of a face or a car, and co-activation of associated representations, i.e. activation of both face and car representations? This question has been teasing me from the hypothesis stage of this project and is still an intense topic of questioning whether inside my mind, with my advisors or during interactions with reviewers. To set the debate it would be useful to lay some premises:

- Neural representations aren't the actual object, they are representations, symbols, more concretely they are the connections set up genetically or through the environment, and most probably both, that link (for example) a retinal stimulation with a neural pattern of activation. So if a representation changes does it mean that the actual object changed? Obviously not, it means that the circuitry that it triggers has been modified in the brain.
- But it appears obvious that some representations can trigger the activation of others. This is actually the definition of associative memory.
- So when and how can we talk about a modulation of a specific representation and when are we talking about two distinct representations activating each other?

It appears obviously extreme to consider that, in the case of perception, any activity pattern evoked by the presentation of, let us say, a class of stimuli such as cars is the neural representation of this class of stimuli. First of all because of top-down processes, which can make this pattern different not due to the sensory inputs but the state of the neural system. For example the predictive coding theory has been tested on perception processes and it has been shown that top-down processes can modulate the multivoxel response patterns in sensory cortices. For instance de Gardelle & al (de Gardelle et al., 2013) showed that different levels of expectation signals could differentially modulate responses of distinct populations of voxels in the FFA in the context of a repeating presentation of faces. They suggest that expectations occurred at different levels: a "higher" level representing the expectations about the sequence of stimuli and attenuating the surprise of seeing a face repeating and a "lower" level representing an accrued information about the stimulus and thus building up over the already present information from the preceding presentation of the same face. Also, considering the shift in neural representation presented in chapter 2, saying that patterns evoked by presentation of faces post-learning has changed in its "core" rather than saying that it is a co-activation, knowing the sluggish temporal resolution of fMRI, is rather extreme and cannot be disentangled from our study. In general the former view can lead to a very relativist account of cognition/neural representations/code, where no representation is discernable as long as there is a temporal or spatial connection with another one, i.e. that would have been triggered by a different external input.

If we push this relativism to the end there would not even be any difference between any mental content or neural representations, which would lead for example to saying that the neural representation of a face = neural representation of a car = neural representation of a voice = *any other cognitive or neural content*.

But there are in my mind some ways to tackle this ambiguity. Let us think about this in terms of neural representations of two distinct visual objects that have been associated, I propose that there are at least two (trivial) ways of taking apart a change in representation from a co-activation:

- Temporally: if the neural representation (e.g. a multi-voxel pattern of activity) elicited by one of two associated visual objects is more similar to its associated pair after than before at the time the representation of the presented object is first represented in the neural activity then it suggests that we have created a new representations / changed the existing ones. For example let us take the example of objects A and B which elicit two different neural representations that both appear at time X after presentation of A or B. If after learning an association between A and B, the neural representation elicited by the presentation of A is more similar to B at time X (or relatively close to time X) then it seems more relevant to talk about a modulation of A's representation or the appearance of a new A-B representation.
- Spatially: if the two stimuli have topographically "distant" neural representations that are both evoked by presenting any of the two associated stimuli after they have been associated, then it seems more reasonable to talk about a co-activation rather than modulation of these representations, which can also happen separately. That is if we accept the premise of a not completely distributed code of visual objects in the brain, in which there could not be any topographically "distant" neural representations in the first place. This "spatial" argument also works in terms of multi-variate representations that are distant in the multi-dimensional space of features, e.g. neurons, voxels, time-points, electrodes.

These suggestions still need to be more specific, and the details will depend on the specific experiment or claims but in general:

- Measure the overlapping of patterns between temporally **distant** moments (pre and post learning) to see if these patterns have changed in terms of location in brain topography.
- Measure the effect of mere exposition to stimuli in order to discard effect of familiarity even in early stages of visual processing.
- Evaluate the temporal dynamics of evoked neural representations, such as what has been done lately in studies investigating visual recognition and invariance using time-resolved techniques (Carlson et al., 2013; Chakravarthi et al., 2014; Isik et al., 2014). This would permit to tell apart the “naïve” processing of visual stimuli (if there is such a thing) and most importantly compare it to the post-learning processing of associated stimuli or, more generally, events.

I believe that making clearer and more direct tests of the “distance” between hypothetically linked representations would help get a better understanding of the neural code and more specifically how it is modulated by experience.

C. Top-down control and monitoring of voluntary learning

One particularity of the experiments carried out in this thesis is the fact that all were voluntary learning tasks. This task context thus poses some demands on top-down processes compared to other types of learning, such as statistical learning of regularities in an environment that can happen outside the subject's consciousness (see Schapiro & Turk-Browne (2015) for a review of these types of learning). The top-down influences from the prefrontal cortices (PFC) have been studied for a long time on many cognitive faculties such as attention or emotions and it has recently been shown that they also exert a very important influence on Medial Temporal Lobe structures and format in which the information is encoded and retrieved from memory (Simons and Spiers, 2003).

Studies investigating the role of MTL structures have been carried on humans and animal models such as rodents and non-human primates. It seems that most of the MTL structures are share some similarity between these species except for

parahippocampal cortex in primates, which is often called postrhinal cortex in rodents, and perirhinal boundaries are less clear in humans than rodents and non-human primates. The frontal cortex, on the other hand, is much more developed in primates than rodents and is roughly similar across primate species in terms of volume relative to the whole brain. Comparative studies suggest that the supposedly higher cognitive functions attributed to human cognition in comparison to great apes might come from a greater connectivity within the PFC areas (Semendeferi et al., 2002) underlying the importance of interactions between areas. Most of what we know about PFC and MTL interactions thus comes from animal models and regarding non-human primates many researchers think that most of it is generalizable to humans.

Anatomically PFC and MTL are connected through large cortico-cortical pathways such as the uncinate fascicle anterior temporal stem and anterior corpus callosum. For example the orbitofrontal and dorsolateral PFC have reciprocal connections with the PRC and ERC. There is also an asymmetry between PFC and PRC in that there are more connections from the PFC to this part of the MTL than reciprocally, suggesting a mainly top-down influence from the PFC (Rempel-Clower and Barbas, 2000).

Historically the role of the PFC has not been often considered in memory processes due to the type of symptoms that are of more behavioral relevance following PFC lesions such as disinhibition and impulsiveness. But it is now clear that damage to certain parts of the PFC can for instance induce memory related impairments such as confabulation which is considered to be an impairment of specification of retrieval task parameters and monitoring and verification of recollected content, i.e. "does this retrieved information fits with the rest of the recollected memory?" (Burgess, 1996). Another neuropsychological study reported the case of an uncinate fascicle lesion in which the patient was particularly impaired at recalling autobiographical events from his own past but was able to perform on tests of new learning (Levine et al., 1998), highlighting the importance of the connections between PFC and MTL.

Voluntary learning, as performed in our studies, can be seen as an encoding process that organizes the incoming information following task instructions or at least requires the subjects to focus on the ordered presentation of items in study 1, or the correct association to make between stimuli through the negative auditory feedback in study 2. For instance in the study presented in chapter 1 different aspects of the paradigm could have been the focus of learning: how many times did the gray square appear relative to the actual images? What was the number of images in the sequence? etc. But the instructions were "learn the sequence" which led to specific strategies to create a "mental map" or an ordered link between the different images.

It has been shown that specific PFC regions are involved in this kind of top-down interaction with the MTL. Studies on non-human primates as well as neuroimaging studies in humans have found evidence that lateral PFC was linked with goal-directed cognitive control functions supporting encoding of discrete memory traces and their retrieval. There are seemingly two important subdivisions in this part of the PFC: a dorsal and ventral part of the lateral PFC. The dorsolateral PFC (dlPFC) is thought to organize the material to be remembered and at the retrieval stage evaluates, monitors and verifies material which has been retrieved. Ventrolateral PFC (vlPFC), on the other hand, seems to underlie elaborative language related functions of memory such as semantic and phonological processing of MTL representations as well as specification of retrieval cues and maintenance of the retrieved information at the retrieval stage, see Figure 26 for a summary of PFC subregions involvement in encoding and retrieval processes.

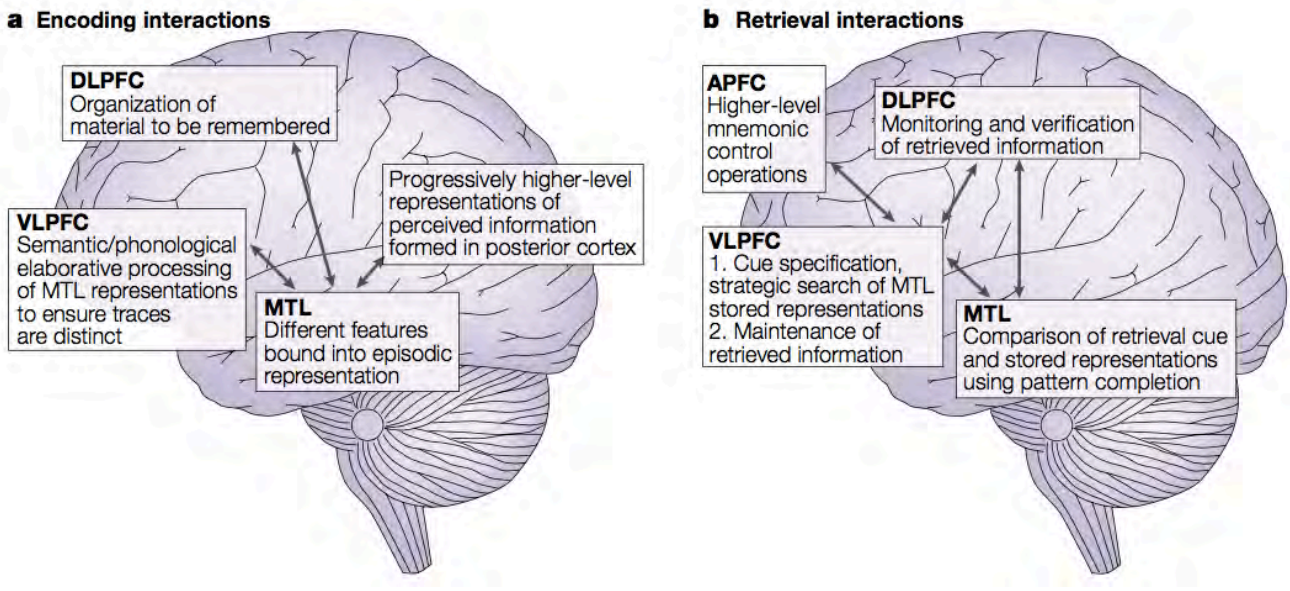

Figure 26. Summary of the roles of different parts of the anterior and lateral PFC in encoding and retrieval processes. from (Simons and Spiers, 2003)

Although the functions that seem to be implemented by these parts of the PFC are difficult to disentangle in a task such as visual-sequence learning, where the subject is constantly required to store and retrieve the information to learn, it is clear that vlPFC and dlPFC must play a crucial role in performing the task. In the same way, the learning protocol designed in the study presented in chapter 2 also involved top-down processes in order to encode the associations and retrieve them, which seem to be carried out by these two regions. Because the scope of this thesis did not encompass the characterization of the top-down processes, most probably carried out by different regions of PFC, we did not aim at recording these regions in the fMRI study and using EEG did not permit to localize precisely where the top-down signals that elicited selective representations in chapter 1's study came from.

It would therefore be crucial to investigate how these regions were involved at the learning and post-learning stages. It would also be very interesting to characterize when each of them was involved, i.e. which part of the lateral PFC was involved along the learning process, and how these PFC parts interact with each other.

More specific to the type of learning used in the work presented in this thesis, some studies have investigated how associative learning manifests in the PFC. A study by Asaad & al (1998) on associative learning mechanisms in the PFC, showed that in a cue saccade association task neurons in the monkey dlPFC seem to receive convergent information and to form associations between a specific cue and the associated saccade direction. Furthermore the protocol was designed to observe the appearance of associative behaviour in neuronal activity by reversing the association between cue and saccade direction, e.g. cue A indicated saccade to the target on the left and then switched to indicate saccade to the target on the right. They observed that learning took place in a few tens of trials: the neurons which were representing cue 1 -> leftward saccade "learned" the new rule and started to represent the opposite association. This result shows that learning behaviour can be observed at the neuronal level in this part of the PFC, which we previously referred to as being a potentially important part of the processes involved in the tasks used in this thesis. The same group led by Earl Miller realized another study later on (Rainer et al., 1999) where they confirmed that the PFC does not only buffer incoming information from different posterior cortex and MTL areas but selectively processes information in a goal-relevant and prospective manner. In this study monkeys performed a delayed paired associate task, between pairs of images, and some neurons located in monkey dlPFC were at first selectively representing the cue and, at the end of the delay, selectively representing the target, which indicated a prospective code. These studies show a flexible associative code implemented in the dlPFC that fits relatively well with the demands of our tasks even though it does not address directly how, during the learning period, the dlPFC or other PFC regions manipulated the content to be associated through direct interaction with inferotemporal cortex or through the MTL.

Understanding how the PFC directs other structures such as the MTL or the VVP during the acquisition of information in the form of associations or else, would give us great insights on what might go wrong when PFC structures or their connections with the rest of the brain regions are lesioned. We also know that developmentally PFC regions don't mature at the same pace and in the same way (Kolb et al., 2012), understanding their role in voluntary learning could help us design better learning programs for children and in cases of developmental

disorders in which PFC functions are impaired either train these functions specifically or cope with these handicaps and adapt programs in order to help them to learn.

Taking some perspective and distance from what we know about how and where the PFC implements memory related processes it appears that these processes themselves are learned at different periods of life: from perinatal, infancy even until adulthood. This idea was coined by Abraham and Bear (1996) and termed "Metaplasticity", which basically represents the idea that one's life experiences affect neural structure and might thus alter plasticity properties, in a positive or negative way. Relative to the PFC many studies have shown that different sub-regions are altered by early life experiences such as exposition to psychoactive drugs, parental care or pre and post-natal stressful events. A direct consequence of this observation/theory is that in the context of learning deficits a promising field of study is to better characterize "normal" PFC functioning and when these functions develops in order to design rehabilitation protocols that would help build these functions in cases where they do not appear to work properly. The findings presented in this thesis can serve as evidence of what "should" happened in specific associative learning contexts, at short and long time scales, and might be of utility when comparing how it takes place in children suffering from learning disabilities.

D. The missing lobe

For the sake of completeness let us talk about the parietal lobe whose contribution to memory has been put forward by Wagner & al (2005) thanks to their review. Consistent parietal lobe activations from the neuroimaging literature in memory tasks has led the field to question the role it played in memory processes and meta-analyses of such studies have highlighted that the parietal lobe seems to be involved in recollection tasks rather than familiarity tasks (Simons et al., 2008). One complicating factor was the fact that patients suffering from parietal lobe lesions did not show any impairment in a recollection task even when their lesions overlapped closely with parietal activation during the same task in healthy subjects. A closer examination of the different aspects of recollection revealed that

their performance might not be completely normal, for example these patients produced less detailed recall and a lack of vividness in freely recalled autobiographical events. This hypothesis was addressed by Simons & al (2009) in a study where they incorporated both measures of objective recollection (accuracy of the recollected events) and subjective measures (trial-by-trial confidence ratings) and were able to show that although there was no objective memory impairment, bilateral parietal lobe lesion patients gave reduced subjective recollection ratings, indicating that their personal experience of recollection may be diminished.

Thus it appears that the parietal lobe's involvement in memory processes are not of particular interest in the work presented in this thesis because it did not deal with the subjectivity of associative learning and recollection but it might be relevant for future studies involving the subjective experience of automatic recollection of associated items in tasks comparable to the one we used. Such question was investigated in a recent study (Leiker and Johnson, 2015). Subjects went through a word-task encoding phase: words were presented while subjects had to perform different tasks on each depending on the instructions. Subjects then saw this set again mixed with new words and had to do a source memory judgment, i.e. in which task did they encounter this word or is it a new word, and rate their confidence level on this response. They first showed that there was a reinstatement of the activity patterns evoked during encoding, which had already been shown in previous studies, and more interestingly that the magnitude of this cortical reinstatement varied with the level of confidence of source memory judgment reported by the subjects. More importantly activity in the left posterior parietal cortex was correlated with the reactivation measure at the trial level. These results thus strengthen the hypothesis that the posterior parietal cortex plays a role in cortical reinstatement of encoded memories and that this reactivation contributes to the subjective aspects of episodic recollection. This study is different from ours in a very important manner which is that subject were voluntarily recalling events, whereas in our tasks we only analyzed neural activity during periods of when subjects were not instructed to recall the learned content. Despite this difference it is clear that during learning our subjects had to recall learned contents and had a feeling of how confident they were about this

knowledge. It would thus be interesting to understand how the parietal cortex reinstatement role and its correlation with confidence measures was implicated in the representational shift observed in our fMRI study presented in chapter 2. For instance repeating the experiment with whole-brain recording would permit to disambiguate whether the patterns of activity evoked by the presentation of a category were modulated by parietal lobe activity, such as in the study reported above, e.g. when presented with faces did the parietal cortex orchestrated the reinstatement of the representation of the associated category, e.g. cars. If this is the case it would strengthen the co-activation interpretation of our results.

As suggested in this study and by a review by Yazar & al (2012) the parietal lobe plays an important role in confidence aspects of the retrieved memories. It might thus also be interesting to see if the magnitude of the shift in neural representation, i.e. difference in similarity between associated categories across sessions, would be correlated with subjects' confidence in their knowledge of the learned associations. To investigate this latter question a event-related design would be more sensitive and thus more relevant because of the differences that might exist between different item pairs.

Other aspects of subjective experience have been linked to memory and perceptual processes such as mental imagery.

E. Mental imagery

The field of mental imagery, and introspection in general, has been through periods of fame and shame, from one of its seminal descriptions by Francis Galton (1907) to its repudiation by behaviorists in the 1930's. But it has been brought back to the central stage with the cognitive "revolution" and thanks to some of its most prominent investigators such as Stephen Kosslyn and is now recognized as an important and sometimes even necessary cognitive faculty.

For instance, it has been suggested that mental imagery had a functional role in associative learning mechanisms (Pearson and Westbrook, 2015). It would thus be interesting to investigate how subjective measures of associative learning and recollection vary according to the performance level in the 2AFC task in chapter

2's experiment or in the level of spontaneous selective activations found in chapter 1's experiment. For example in both studies some subjects reported having some spontaneous mental images of the associated or expected stimulus. Because these reports were not planned in the experimental design and subjective mental imagery was not the focus of these studies it was not possible to correlate these with individual differences in neural manifestations of learning effects.

This review argued that a dichotomy between voluntary and involuntary invoked mental imagery is relevant to better understand what constitutes conscious sensory perception. Indeed the results of both experiments presented in this thesis could be counted as feats of mental imagery and one possible development following the work presented here would be to better characterize how mental imagery is evoked and how to disentangle mental imagery from automatic activation of a absent visual stimulus' neural representation or changes in this specific neural representation, or at least consider if such a dissociation is possible and/or relevant.

Conclusion

The work realized in this thesis has, modestly, extended the current knowledge on the flexibility of neural processing and neural representations in multiple aspects. We characterized the first steps of sequential visual learning in terms of neural mechanisms involved and the timing of these processes which informs us on how we create rapid links between events occurring in a fixed temporal order and that could be linked with other cognitive abilities such as spatial navigation. We were also able to show a very surprising plasticity of the ventral visual pathway's existing and well-established neural representations of visual objects. We were also able to link this modulation with markers of perception, such as categorization abilities, that should be the final endeavor for any cognitive neuroscience investigation in order to bridge the gap between fundamental research and application in improving learning programs, better understanding learning disabilities and designing rehabilitation programs. For example understanding the mechanisms that shape the VVP and how it connects to perceptual abilities can help us understand how sometimes visual learning does not take place correctly, for instance in dyslexia, or how to better tailor reading acquisition for adults and monitor teaching programs to have an objective measure of integration of different kinds of information.

References

Abraham, W.C., and Bear, M.F. (1996). Metaplasticity: the plasticity of synaptic plasticity. Trends Neurosci. *19*, 126–130.

Arthurs, O.J., and Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? Trends Neurosci. *25*, 27–31.

Asaad, W.F., Rainer, G., and Miller, E.K. (1998). Neural Activity in the Primate Prefrontal Cortex during Associative Learning. Neuron *21*, 1399–1407.

Baizer, J.S., Ungerleider, L.G., and Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J. Neurosci. *11*, 168–190.

Beeck, H.P. Op de, and Baker, C.I. (2010). The neural basis of visual object learning. Trends Cogn. Sci. *14*, 22–30.

Beeck, H.P. Op de, Baker, C.I., DiCarlo, J.J., and Kanwisher, N.G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. J. Neurosci. *26*, 13025–13036.

Beeck, H.P. Op de, Haushofer, J., and Kanwisher, N.G. (2008). Interpreting fMRI data: maps, modules and dimensions. Nat. Rev. Neurosci. *9*, 123–135.

Benson D, and Greenberg JP (1969). Visual form agnosia: A specific defect in visual discrimination. Arch. Neurol. *20*, 82–89.

Berger, P.D.H. (1929). Über das Elektrenkephalogramm des Menschen. Arch. Für Psychiatr. Nervenkrankh. *87*, 527–570.

Bliss, T.V.P., and Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J. Physiol. *232*, 331–356.

Burgess, P.W. (1996). Confabulation and the Control of Recollection. Memory *4*, 359–412.

Buzsaki, G. (2006). Rhythms of the Brain (Oxford University Press, USA).

Carlson, T., Tovar, D.A., Alink, A., and Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. J. Vis. *13*, 1–1.

Chadwick, M.J., Hassabis, D., Weiskopf, N., and Maguire, E.A. (2010). Decoding Individual Episodic Memory Traces in the Human Hippocampus. Curr. Biol. *20*, 544–547.

Chakravarthi, R., Carlson, T.A., Chaffin, J., Turret, J., and VanRullen, R. (2014). The Temporal Evolution of Coarse Location Coding of Objects: Evidence for Feedback. J. Cogn. Neurosci. *26*, 2370–2384.

Charest, I., Kievit, R.A., Schmitz, T.W., Deca, D., and Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. Proc. Natl. Acad. Sci. *111*, 14565–14570.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., and Michel, F. (2000). The visual word form area. Brain *123*, 291–307.

Conner, I., Sharma, S., Lemieux, S., and Jd, M. (2004). Retinotopic organization in

children measured with fMRI. J. Vis. *4*, 509–523.
David, S.V., Hayden, B.Y., and Gallant, J.L. (2006). Spectral Receptive Field Properties Explain Shape Selectivity in Area V4. J. Neurophysiol. *96*, 3492–3505.
Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Filho, G.N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., and Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. Science *330*, 1359–1364.
Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. Nat. Rev. Neurosci. *16*, 234–244.
Desimone, R., Albright, T.D., Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci. *4*, 2051–2062.
Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. Science *293*, 2470–2473.
Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. Nat. Rev. Neurosci. *15*, 732–744.
Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598–601.
Van Essen, D. (2004). Organization of Visual Areas in Macaque and Human Cerebral Cortex. In Visual Neuroscience, L.M. Chalupa, and J.S. Werner, eds. (MIT Press), pp. pp. 507–521.
Farah, M.J. (1992). Is an Object an Object an Object? Cognitive and Neuropsychological Investigations of Domain Specificity in Visual Object Recognition. Curr. Dir. Psychol. Sci. *1*, 164–169.
Farah, M.J., Hammond, K.M., Levine, D.N., and Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. Cognit. Psychol. *20*, 439–462.
Feinberg, T.E., and Farah, M.J. (2000). Patient-based Approaches to Cognitive Neuroscience (MIT Press).
Felleman, D.J., and Essen, D.C.V. (1991). Distributed Hierarchical Processing in the Primate. Cereb. Cortex *1*, 1–47.
Földiák, P. (1991). Learning Invariance from Transformation Sequences. Neural Comput. *3*, 194–200.
Foster, D.J., and Knierim, J.J. (2012). Sequence learning and the role of the hippocampus in rodent navigation. Curr. Opin. Neurobiol. *22*, 294–300.
Frankland, P.W., and Bontempi, B. (2005). The organization of recent and remote memories. Nat. Rev. Neurosci. *6*, 119–130.
Fuentemilla, L., Penny, W.D., Cashdollar, N., Bunzeck, N., and Düzel, E. (2010). Theta-Coupled Periodic Replay in Working Memory. Curr. Biol. *20*, 606–612.
Galton, F. (1907). Mental imagery. In Inquiries into Human Faculty and Its Development, (London, England: J M Dent & Sons), pp. 57–79.
de Gardelle, V., Stokes, M., Johnen, V.M., Wyart, V., and Summerfield, C. (2013). Overlapping multivoxel patterns for two levels of visual expectation. Front. Hum. Neurosci. *7*.
Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., and Gore, J.C. (1999). Activation

of the middle fusiform “face area” increases with expertise in recognizing novel objects. Nat. Neurosci. *2*, 568–573.
Gauthier, I., Skudlarski, P., Gore, J.C., and Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. *3*, 191–197.
Gazzaley, A., and D’Esposito, M. (2007). Unifying prefrontal cortex function: Executive control, neural networks, and top-down modulation. In The Human Frontal Lobes: Functions and Disorders (2nd Ed.), B.L. Miller, and J.L. Cummings, eds. (New York, NY, US: Guilford Press), pp. 187–206.
Gazzaley, A., and Nobre, A.C. (2012). Top-down modulation: bridging selective attention and working memory. Trends Cogn. Sci. *16*, 129–135.
Girshick, A.R., Landy, M.S., and Simoncelli, E.P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat. Neurosci. *14*, 926–932.
Golarai, G., Ghahremani, D.G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J.L., Gabrieli, J.D.E., and Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. Nat. Neurosci. *10*, 512–522.
Goodale, M.A., Meenan, J.P., Bülthoff, H.H., Nicolle, D.A., Murphy, K.J., and Racicot, C.I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. Curr. Biol. *4*, 604–610.
Grandchamp, R. (2012). Electroencéphalographie et interfaces cerveau-machine : nouvelles méthodes pour étudier les états mentaux. phd. Université de Toulouse, Université Toulouse III - Paul Sabatier.
Grech, R., Cassar, T., Muscat, J., Camilleri, K.P., Fabri, S.G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., and Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. J. NeuroEngineering Rehabil. *5*, 1–33.
Grill-Spector, K., and Weiner, K.S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. *15*, 536–548.
Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., and Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum. Brain Mapp. *6*, 316–328.
Grill-Spector, K., Golarai, G., and Gabrieli, J. (2008). Developmental neuroimaging of the human ventral visual cortex. Trends Cogn. Sci. *12*, 152–162.
Gross, C.G., Rocha-Miranda, C.E., and Bender, D.B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. J. Neurophysiol. *35*, 96–111.
de Haan, M., Pascalis, O., and Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. J. Cogn. Neurosci. *14*, 199–209.
Harrison, S.A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. Nature *458*, 632–635.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science *293*, 2425–2430.
Haxby, J.V., Connolly, A.C., and Guntupalli, J.S. (2014). Decoding neural

representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. *37*, 435–456.
Haynes, J.-D., and Rees, G. (2005). Predicting the Stream of Consciousness from Activity in Human Visual Cortex. Curr. Biol. *15*, 1301–1307.
Hebb, D.O. (2005). The Organization of Behavior: A Neuropsychological Theory (Psychology Press).
Hubel, D.H., and Wiesel, T.N. (1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. *148*, 574–591.
Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol *160*, 106–154.
Isik, L., Leibo, J.Z., and Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. Front. Comput. Neurosci. *6*.
Isik, L., Meyers, E.M., Leibo, J.Z., and Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. J. Neurophysiol. *111*, 91–102.
Izquierdo, I., and Medina, J.H. (1997). Memory Formation: The Sequence of Biochemical Events in the Hippocampus and Its Connection to Activity in Other Brain Structures. Neurobiol. Learn. Mem. *68*, 285–316.
Jensen, O., and Lisman, J.E. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. Trends Neurosci. *28*, 67–72.
Jones, J.P., and Palmer, L.A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. *58*, 1233–1258.
Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.
Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. J. Neurosci. *17*, 4302–4311.
Kaplan, J.T., Man, K., and Greening, S.G. (2015). Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. Front. Hum. Neurosci. 151.
Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.
Kobatake, E., and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J. Neurophysiol. *71*, 856–867.
Kok, P., Failing, M.F., and de Lange, F.P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. J. Cogn. Neurosci. *26*, 1546–1554.
Kolb, B., Mychasiuk, R., Muhammad, A., Li, Y., Frost, D.O., and Gibb, R. (2012). Experience and the developing prefrontal cortex. Proc. Natl. Acad. Sci. *109*, 17186–17193.
Kosslyn, S.M., Thompson, W.L., and Ganis, G. (2006). The case for mental imagery (New York, NY, US: Oxford University Press).
Kourtzi, Z., and Connor, C.E. (2011). Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. Annu. Rev. Neurosci. *34*, 45–67.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008a). Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience. Front. Syst. Neurosci. *2*.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008b). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron *60*, 1126–1141.
Leiker, E.K., and Johnson, J.D. (2015). Pattern reactivation co-varies with activity in the core recollection network during source memory. Neuropsychologia *75*, 88–98.
Levine, B., Black, S.E., Cabeza, R., Sinden, M., Mcintosh, A.R., Toth, J.P., Tulving, E., and Stuss, D.T. (1998). Episodic memory and the self in a case of isolated retrograde amnesia. Brain J. Neurol. *121 (Pt 10)*, 1951–1973.
Lisman, J., and Redish, A.D. (2009). Prediction, sequences and the hippocampus. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 1193–1201.
Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150–157.
Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., and Tootell, R.B.H. (1995). Object-Related Activity Revealed by Functional Magnetic Resonance Imaging in Human Occipital Cortex. Proc. Natl. Acad. Sci. U. S. A. *92*, 8135–8139.
Malmivuo, J., and Plonsey, R. (1995). Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields (Oxford University Press).
Manns, J.R., Howard, M.W., and Eichenbaum, H. (2007). Gradual Changes in Hippocampal Activity Support Remembering the Order of Events. Neuron *56*, 530–540.
Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells$\ast$. J Opt Soc Am *70*, 1297–1300.
Mayes, A., Montaldi, D., and Migo, E. (2007). Associative memory and the medial temporal lobes. Trends Cogn. Sci. *11*, 126–135.
McCandliss, B.D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn. Sci. *7*, 293–299.
McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. *102*, 419–457.
Mehta, M.R., Barnes, C.A., and McNaughton, B.L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. Proc. Natl. Acad. Sci. *94*, 8918–8921.
Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. Nature *273*, 297–298.
Mishkin, M., and Ungerleider, L.G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav. Brain Res. *6*, 57–77.
Mishkin, M., Ungerleider, L.G., and Macko, K.A. (1983). Object vision and spatial vision: two cortical pathways. Trends Neurosci. *6*, 414–417.

Mitchell, D.J., McNaughton, N., Flanagan, D., and Kirk, I.J. (2008). Frontal-midline theta from the perspective of hippocampal "theta." Prog. Neurobiol. *86*, 156–185.
Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature *335*, 817–820.
Miyashita, Y., Kameyama, M., Hasegawa, I., and Fukushima, T. (1998). Consolidation of Visual Associative Long-Term Memory in the Temporal Cortex of Primates. Neurobiol. Learn. Mem. *70*, 197–211.
Mondloch, C.J., Geldart, S., Maurer, D., and Le Grand, R. (2003). Developmental Changes in Face Processing Skills. J. Exp. Child Psychol. *86*, 67–84.
Moser, E.I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. Annu. Rev. Neurosci. *31*, 69–89.
Nowak, L.G., and Bullier, J. (1997). The Timing of Information Transfer in the Visual System. In Extrastriate Cortex in Primates, K.S. Rockland, J.H. Kaas, and A. Peters, eds. (Springer US), pp. 205–241.
Ogawa, S., Lee, T.M., Kay, A.R., and Tank, D.W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. *87*, 9868–9872.
O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. *34*, 171–175.
Olavarria, J.F., and Essen, D.C.V. (1997). The global pattern of cytochrome oxidase stripes in visual area V2 of the macaque monkey. Cereb. Cortex *7*, 395–404.
Olejniczak, P. (2006). Neurophysiologic basis of EEG. J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc. *23*, 186–189.
Pasupathy, A., and Connor, C.E. (2002). Population coding of shape in area V4. Nat Neurosci *5*, 1332–1338.
Pearson, J., and Westbrook, F. (2015). Phantom perception: voluntary and involuntary nonretinal vision. Trends Cogn. Sci. *19*, 278–284.
Pegado, F., Comerlato, E., Ventura, F., Jobert, A., Nakamura, K., Buiatti, M., Ventura, P., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., et al. (2014). Timing the impact of literacy on visual processing. Proc. Natl. Acad. Sci. U. S. A. *111*, E5233–E5242.
Perry, G., Rolls, E.T., and Stringer, S.M. (2010). Continuous transformation learning of translation invariant representations. Exp. Brain Res. *204*, 255–270.
Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? Trends Cogn. Sci. *7*, 113–118.
Rainer, G., Rao, S.C., and Miller, E.K. (1999). Prospective Coding for Objects in Primate Prefrontal Cortex. J. Neurosci. *19*, 5493–5505.
Reddy, L., Tsuchiya, N., and Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. NeuroImage *50*, 818–825.
Reddy, L., Poncet, M., Self, M.W., Peters, J.C., Douw, L., van Dellen, E., Claus, S., Reijneveld, J.C., Baayen, J.C., and Roelfsema, P.R. (2015). Learning of anticipatory responses in single neurons of the human medial temporal lobe. Nat. Commun. *6*, 8556.
Rempel-Clower, N.L., and Barbas, H. (2000). The laminar pattern of connections

between prefrontal and anterior temporal cortices in the Rhesus monkey is related to cortical structure and function. Cereb. Cortex N. Y. N 1991 *10*, 851–865.
Rossetti, Y., Pisella, L., and Vighetto, A. (2003). Optic ataxia revisited. Exp. Brain Res. *153*, 171–179.
Sakai, K., and Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. Nature *354*, 152–155.
Schapiro, A., and Turk-Browne, N. (2015). Statistical Learning. In Brain Mapping, A.W. Toga, ed. (Waltham: Academic Press), pp. 501–506.
Schapiro, A.C., Kustner, L.V., and Turk-Browne, N.B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. Curr. Biol. *22*, 1622–1627.
Schneider, G.E. (1969). Two visual systems. Science *163*, 895–902.
Scoville, W.B., and Milner, B. (1957). LOSS OF RECENT MEMORY AFTER BILATERAL HIPPOCAMPAL LESIONS. J. Neurol. Neurosurg. Psychiatry *20*, 11–21.
Semendeferi, K., Lu, A., Schenker, N., and Damasio, H. (2002). Humans and great apes share a large frontal cortex. Nat. Neurosci. *5*, 272–276.
Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust Object Recognition with Cortex-Like Mechanisms. IEEE Trans. Pattern Anal. Mach. Intell. *29*, 411–426.
Simons, J.S., and Spiers, H.J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. Nat. Rev. Neurosci. *4*, 637–648.
Simons, J.S., Peers, P.V., Hwang, D.Y., Ally, B.A., Fletcher, P.C., and Budson, A.E. (2008). Is the parietal lobe necessary for recollection in humans? Neuropsychologia *46*, 1185–1191.
Simons, J.S., Peers, P.V., Mazuz, Y.S., Berryhill, M.E., and Olson, I.R. (2009). Dissociation Between Memory Accuracy and Memory Confidence Following Bilateral Parietal Lesions. Cereb. Cortex bhp116.
Smith, E.E., and Kosslyn, S.M. (2007). Encoding and retrieval from long-term memory. Cogn. Psychol. Mind Brain 195–246.
Squire, L.R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. Science *253*, 1380–1386.
Srebro, B. (2012). Looking back at Jerzy Konorski's book "Integrative Activity of the Brain" 45 years after. Acta Neurobiol. Exp. (Warsz.) *73*, 451–462.
Summerfield, C., Jack, A.I., and Burgess, A.P. (2002). Induced gamma activity is associated with conscious awareness of pattern masked nouns. Int. J. Psychophysiol. *44*, 93–100.
Suzuki, W.A., and Amaral, D.G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. J. Comp. Neurol. *350*, 497–533.
Suzuki, S., Augerinos, G., and Black, A.H. (1980). Stimulus control of spatial behavior on the eight-arm maze in rats. Learn. Motiv. *11*, 1–18.
Takashima, A., Petersson, K.M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M.J., McNaughton, B.L., and Fernández, G. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. Proc. Natl.

Acad. Sci. U. S. A. *103*, 756–761.
Tanaka, K. (1996). Inferotemporal Cortex and Object Vision. Annu. Rev. Neurosci. *19*, 109–139.
Tangen, G.G., Engedal, K., Bergland, A., Moger, T.A., Hansson, O., and Mengshoel, A.M. (2015). Spatial navigation measured by the Floor Maze Test in patients with subjective cognitive impairment, mild cognitive impairment, and mild Alzheimer's disease. Int. Psychogeriatr. *27*, 1401–1409.
Tarr, M.J., and Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat. Neurosci. *3*, 764–769.
Thorpe, S.J., and Fabre-Thorpe, M. (2001). Seeking Categories in the Brain. Science *291*, 260–263.
Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature *381*, 520–522.
Turnbull, O.H., Carey, D.P., and McCARTHY, R.A. (1997). The neuropsychology of object constancy. J. Int. Neuropsychol. Soc. *3*, 288–298.
VanRullen, R., and Thorpe, S.J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. Perception *30*, 655–668.
VanRullen, R., Delorme, A., and Thorpe, S. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. Neurocomputing *38–40*, 1003–1009.
Vidal, J.R., Ossandón, T., Jerbi, K., Dalal, S.S., Minotti, L., Ryvlin, P., Kahane, P., and Lachaux, J.-P. (2010). Category-specific visual responses: an intracranial study comparing gamma, beta, alpha, and ERP response selectivity. Front. Hum. Neurosci. *4*, 195.
Vuilleumier, P., Henson, R.N., Driver, J., and Dolan, R.J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat. Neurosci. *5*, 491–499.
Wagner, A.D., Shannon, B.J., Kahn, I., and Buckner, R.L. (2005). Parietal lobe contributions to episodic memory retrieval. Trends Cogn. Sci. *9*, 445–453.
Wallenstein, G.V., Hasselmo, M.E., and Eichenbaum, H. (1998). The hippocampus as an associator of discontiguous events. Trends Neurosci. *21*, 317–323.
Wandell, B.A., and Winawer, J. (2015). Computational neuroimaging and population receptive fields. Trends Cogn. Sci. *19*, 349–357.
Wilson, M.A., and McNaughton, B.L. (1993). Dynamics of the hippocampal ensemble code for space. Science *261*, 1055–1058.
Wolbers, T., and Hegarty, M. (2010). What determines our navigational abilities? Trends Cogn. Sci. *14*, 138–146.
Yazar, Y., Bergström, Z.M., and Simons, J.S. (2012). What is the parietal lobe contribution to long-term memory? Cortex J. Devoted Study Nerv. Syst. Behav. *48*, 1381–1382; discussion 1383–1387.
Zeki, S. (2003). The disunity of consciousness. Trends Cogn. Sci. *7*, 214–218.
Zeki, S.M. (1969). Representation of central visual fields in prestriate cortex of monkey. Brain Res. *14*, 271–291.

Zola-Morgan, S.M., and Squire, L.R. (1990). The primate hippocampal formation: evidence for a time-limited role in memory storage. Science *250*, 288–290.
(2015). Place cell.
H. Lissauer (1890). Ein Fall von Seelenblindheit nebst einem beitrage zue Theorie derselben. Archiv fur Psychiatrie und Nervenkrankheiten.

AUTEUR : Mehdi Senoussi

TITRE :
Temporal and spatial flexibility of neural representations of visual objects through learning

Flexibilité temporelle et spatiale des représentations neurales d'objets visuels lors d'apprentissages

DIRECTEURS DE THESE : Rufin VanRullen et Leila Reddy

LIEU ET DATE DE SOUTENANCE :
Salle de conférence du Pavillong Baudot
CNRS CERCO UMR 5549
Pavillon Baudot CHU Purpan
BP 25202
31052 Toulouse Cedex

Le Vendredi 5 février 2016

RÉSUMÉ EN FRANÇAIS :
Les travaux présentés dans cette thèse portent sur l'effet d'apprentissages à court et long terme sur le système visuel. Nous avons d'abord montré grâce à des enregistrements Éléctroencéphalographiques que l'apprentissage d'une séquence de stimuli visuels induisait une activité cérébrale spontanée et sélective au prochain stimulus devant apparaitre et que cette activité sélective s'exprimait dans les bandes alpha et beta de l'activité électrique cérébrale. Par la suite nous avons montré grâce à de l'Imagerie par Résonance Magnétique fonctionnelle que lors d'apprentissages longs (trois semaines) les représentations neurales de catégories visuelles associées étaient modulées et devenaient plus similaires après l'apprentissage. Les travaux présentés dans cette thèse ont donc permis de mieux caractériser l'impact d'apprentissages à différentes échelles de temps sur les représentations neurales d'objets visuels.

MOTS CLÉS :
Apprentissages, représentations neurales, objets visuels, IRM fonctionnelle, EEG

Learning, neural representations, visual objects, functional MRI, EEG

DISCIPLINE :
Neurosciences, comportement et cognition

LABORATOIRE :
CNRS CERCO UMR 5549
Pavillon Baudot CHU Purpan
BP 25202
31052 Toulouse Cedex